BARRIERS TO GENE EXCHANGE IN A FIELD CRICKET HYBRID ZONE.

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Luana S. Maroja

August 2008

BARRIERS TO GENE EXCHANGE IN A CRICKET HYBRID ZONE.

Luana S. Maroja, Ph. D.

Cornell University 2008

Understanding how speciation influences patterns of molecular evolution and how molecular changes drive speciation are central questions in evolutionary biology. In this dissertation I address questions of behavioral barriers to gene exchange, cytoplasmic incompatibility, molecular evolution, population history, gene flow and species boundaries in the hybridizing field crickets *Gryllus firmus* and *G. pennsylvanicus*. I examine the role of behavioral barriers to gene exchange in the context of previous studies that documented temporal and ecological isolation and a one-way post-mating incompatibility. My results reveal strong behavioral premating barriers, but no apparent fecundity or fertility costs for *G. firmus* females when they mate with conspecific and heterospecific males. I also document a failure of heterospecific males to induce normal oviposition in *G. firmus* females, a previously unknown post-mating, pre-zygotic barrier. *Gryllus firmus* and *G. pennsylvanicus* exhibit a very clear unidirectional incompatibility and have been cited as a possible example of *Wolbachia*-induced cytoplasmic incompatibility. *Wolbachia* are cytoplasmically inherited alpha-proteobacteria that can cause cytoplasmic incompatibility in insects. I conduct curing experiments, intra- and interspecific crosses, cytological examination of *Wolbachia* in testes, and *Wolbachia* quantifications via Real-Time PCR. All of the data strongly suggest that *Wolbachia* are not involved in the reproductive incompatibility between *G. firmus* and *G. pennsylvanicus*. Finally I analyze DNA sequence

divergence for seminal protein loci, housekeeping loci, and mtDNA, using a combination of analytical approaches and extensive sampling. In recently diverged species, such as *G. firmus* and *G. pennsylvanicus*, ancestral polymorphism and introgression can cause incongruence between gene trees and species trees. In the face of hybridization only genomic regions that cannot cross the species boundaries will show reciprocal monophyly. These regions, usually evolving rapidly under selection, are essential for the maintenance of species identity. I report discordant genealogical patterns and differential introgression rates across the genome. The most dramatic outliers, showing near zero introgression and more structured species trees, are also the only two seminal protein loci under selection. These are candidate barrier genes with possible reproductive functions. I also use the genealogical data to examine the demographic history and the current structure of the hybrid zone.

BIOGRAPHICAL SKETCH

Luana Santoro Maroja was born on the 19th of October 1976 in Rio de Janeiro, Brazil. Growing up in the tropical Atlantic rain forest, Luana has always been fascinated with Biology. As a young child, she dreamed of becoming a "dinosaur hunter" and arranged many "archeology expeditions" to her backyard pet cemetery where she dug for fossils (to her frustration she was never able to fully assemble any skeleton). Despite the hopes of her parents to go to medical school, Luana decided at 11 years old that her career would follow the path of an academic researching ecology and conservation. After her father's early death, Luana and her family passed through emotional and financial hard times, but despite these difficulties, she earned her way into the Federal University of Rio de Janeiro, Brazil where she began studying Ecology in 1995.

Her first research experience was with Dr. Rui Cerqueira's"Vertebrates Lab," where she worked with population dynamics of small mammals for four years. As an undergraduate, Luana excelled in tattooing opossum's tails and taxiderming all kinds of animals. During this time, she wrote her first publications on small mammal population dynamics and parasitology of a water rat. However, Luana's interests in biology were very broad, spanning evolution, population genetics, ecology, animal behavior and conservation biology. Thus in her senior year, Luana decided to concentrate on evolution and genetics, leading her to pursue a master degree in population genetics of a tropical water rat in Dr. Hector Seuanez's lab. As Luana's interests in evolution grew, she became intellectually attracted to the speciation process and how biodiversity is generated. She decided to pursue a doctoral degree and, in 2001, was fortunate to find the ideal advisor and lab to pursue her growing interests in speciation, when she joined Dr. Richard Harrison's lab in Cornell

University. Luana really valued her time as a graduate student. She not only learned invaluable research on evolutionary biology but also how to be an effective teacher, after being a teaching assistant for six years at Cornell.

Luana truly loved her time in Ithaca. During her Ph.D. studies, Luana met her husband and father of her daughters, John Carasone. Their first daughter, Alani Maroja Carasone, was born on February 02, 2006 and their second daughter, Adriana Maroja Carasone, was born on November 02, 2008. The birth of Luana's daughters made her realize how crucial the role of the educator is and increased her passion for teaching. Luana sees the birth of her daughters not as a hindrance to her career but actually as an inspiration, and as a result she was able to reconcile the increased demands on her time with ease.

For the future, Luana hopes to become a professor at a research institution where she can combine her desire to do biological research with steady teaching. After finishing at Cornell, she and her family will leave to the University of Cambridge where she will begin her first postdoctoral position.

"To Life, The Universe and Everything"

ACKNOWLEDGMENTS

This dissertation is the result of seven years of work and would never have been completed without the help and support of so many talented people. It is impossible to properly acknowledge everyone I would like to with such limited space, but I will start this abbreviated list with those who contributed directly to my growth as a scholar.

I would like to offer very special thanks to my advisor Rick Harrison, who has truly been an ideal advisor. He has sharpened my scientific thinking, is always available to talk (science or non-science topics), makes amazing improvements to written work, encourages students to have a balanced life and fosters great chemistry within the lab group by hosting wonderful parties at his house with his extraordinary wife Ellen. Rick will always serve as an ideal example for me and there could be no better compliment to me if my future students hold me in the same regards that I hold him.

The co-author of most of my work, José Andrés, has been both a great scientist and friend. José cheers me up with his not so cheerful way, and makes lunches and lab meetings a lot more fun. Through him, I've also learned a great deal about statistics and wine (I now know there actually are better wines out there beyond my previous favorite, "Bully Hill"). I hope we will keep in touch for future collaborations and friendship.

Thanks to all my committee members, Charles Aquadro, Carlos Bustamante, Rasmus Nielsen and Kelly Zamudio for helpful insight, great suggestions and advice on statistical analysis.

Thanks to Stephen Bogdanowicz, who somehow squeezes joy out of the monotony of lab work and his extensive knowledge of "all things lab" would make

everyone jealous. In the Harrison lab, there is no greater truism than "without Stevo there would be no results". Rebecca Wilbanks has helped me greatly with sequencing; her sacrifice saved me many hours of boredom. Finally I thank my entire lab group; they have been amazing friends and collaborators: Ben, Erik, Erica, Findley, Jamie, Lee, Leila, Marie, Mike, Mo, Paul, Ryan, Sean and Shoula. I also thank my master degree advisor Dr. Hector Seuanez for helping me prepare for my PhD work and my good friends in Brazil, Cibele, Chica, Miguel, Paulo, Pavel, and Rui all of which contributed in some way for me being here.

Next I thank all people that made my life in Ithaca even more fun. Starting with friends and office mates in the EEB department: Andrea, Angie, Anna, Gretchen, Jason, Jeanne, Jon, Josh (NBB), Krystal, Kurt, Rulon, Sarah among many others. Thanks to my TKD group for teaching me self defense and helping me rise to the level of black belt. Thanks to all the Hasbrouck staff and GCAs for making housing more fun. Thanks to Viviana for adopting and taking good care of my dear rabbit, Branca. Very special thanks to Carmela for being a great friend and teacher to my lovely daughters, Alani and Adriana.

Finally, I deeply thank my family. My husband John has been the best, helping me with the kids, editing English, dealing with countless bureaucracies, keeping finances, making tasty dinners and overall making my life great. My daughters, Alani and Adriana, added spark to my life and it has been a great joy to see them grow. My mother and Tom have been extremely helpful, always coming to help with the kids when we most needed them, while providing great friendship. Tom has also become an honorary evolutionary biologist, developing programs to format genetic data across a multitude of software.

My research was funded by National Science Foundation grants to Richard Harrison (DEB-0639904) and to Willie Swanson (DEB-0111613). I also received 6

years of TAship support from the Cornell Graduate School and EEB department as well as an American Fellowship from the American Association of University Woman (AAUW).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Understanding how speciation influences patterns of molecular evolution and how molecular changes drive speciation are central questions in modern evolutionary biology. In this dissertation I address questions of behavioral barriers to gene exchange, cytoplasmic incompatibility, molecular evolution, population history, gene flow and species boundaries in the hybridizing field crickets *Gryllus firmus* and *G. pennsylvanicus*. Using a combination of behavioral and genetic data, my goal is to understand the maintenance of species barriers and to identify selective and demographic mechanisms that have played a role in speciation.

Understanding reproductive barriers is essential to understanding the maintenance of species boundaries in sexually reproducing organisms and thus why organisms fall into discrete clusters (i.e. species). Reproductive barriers have been traditionally classified as pre- and post-zygotic depending on when they act in the life of an organism.

Pre and post zygotic barrier to gene exchange

Prezygotic barriers to gene exchange include ecological, temporal, behavioral and mechanical barriers that "prevent interspecific crosses" (Mayr, 1963). Prezygotic barriers also include conspecific sperm precedence and cryptic female choice, in which sperm (pollen) transfer occurs but heterospecific sperm (pollen) does not effect fertilization (Mayr, 1963; Coyne and Orr, 2004). Postzygotic barriers include hybrid inviability, infertility and F2 breakdown (Dobzhansky, 1937; Mayr, 1963), as well as ecological inviability and behavioral sterility (Coyne and Orr, 2004).

In many taxa multiple reproductive barriers contribute to isolation (Coyne, 1992; Schluter, 2001; Price and Bouvier, 2002; Ramsey et. al., 2003), but the relative

contribution of each barrier, as well as their importance in the speciation process often remain unknown. Thus to fully understand the evolution of barriers to gene exchange the full set of isolating barriers has to be relatively well understood.

Over the past decade there has been tremendous progress linking genetic data to patterns of isolation between species. Many genes responsible for postzygotic isolation have been described. For example *Xmrk-2* causes inviability in hybrid platyfishes (Wittbrodt *et al.* 1999), *OdsH* and *JYAlpha* cause hybrid male sterility in *Drosophila (*Ting *et al.* 1998; Wu and Ting 2004; Masly *et al.* 2006) and *Hmr, Nup96* and *Lhr* cause hybrid inviability in *Drosophila* (Barbash *et al.* 2003; Presgraves *et al.* 2003; Brideau *et al.* 2006). Most of these genes are rapidly evolving under positive natural selection (Ting *et al.* 1998; Barbash *et al.* 2003; Presgraves *et al.* 2003; Brideau *et al.* 2006), supporting the traditional view that reproductive isolation evolves as an epiphenomenon of Darwinian adaptation (Orr *et al.* 2007). In organisms with more limited genetic resources, linking genes to patterns of isolation between species has been more challenging. Yet, candidate barrier genes have been identified using statistical analysis of hybrid zones (e.g. Riesenberg *et al.* 1999; Grahame *et al.* 2006), analysis of gene genealogies (e.g. Dopman *et al.* 2005; Andres *et al.* submitted), population genetics (e.g. Vasemagi *et al.* 2005; Nosil *et al.* 2008) and coalescent based approaches (e.g. Putnam *et al.* 2007).

Searching for barriers to gene exchange

Many barriers to gene exchange can be involved in the maintenance of species boundaries. Traditionally, postzygotic isolation mechanisms, such as hybrid inviability and infertility, have received the most attention (Harrison 1983; Bert *et al.* 1993; Burke *et al.* 1998; Burke and Arnold 2001). However, attention has increasingly been placed on prezygotic isolation mechanisms, such as conspecific sperm precedence, cryptic female

choice and ecological and behavioral barriers to gene exchange (Cruzan and Arnold 1994;Sætre *et al.* 1997; Howard *et al.* 1998; MacCallum *et al.* 1998; Rieseberg *et al.* 1998; Bailey *et al.* 2004). Behavioral premating barriers are thought to be particularly important because they act early in the life cycle, and thus have the potential to reduce gene flow proportionally more than barriers that act later (Coyne and Orr 2004). Within hybrid zones, pre- and postzygotic barriers are equally important in unimodal zones (those in which hybrid genotypes are common), whereas bimodal hybrid zones (those in which parental types predominate) "are invariably coupled with strong assortative mating or assortative fertilization" (Jiggins and Mallet 2000). Opportunely premating barriers can often be verified in the laboratory with simple mating experiments between conspecific and heterospecific individuals. To find actual genes involved in barriers to gene exchange (i.e. barrier or speciation genes) other approaches must be taken.

Gene genealogies of alleles or haplotypes (gene phylogenies) which are useful tools to identify demographic and historic factors associated with speciation (Hudson 1990; Kliman *et al.* 2000) can also aid in the identification of barrier genes. Because of random sorting of ancestral polymorphism, hybridization leading to introgression, and natural selection (Neigel and Avise 1986; Hudson 1992; Nichols 2001), gene trees and species trees are not always congruent. Recently diverged species have a higher probability of having incongruent gene trees and species trees (Wang *et al.* 1997; Ting *et al.* 2000). If the coalescence time for samples of alleles is more than or equal to the time back to a lineage-splitting event, then most gene trees will not reflect the species tree. Thus, recently diverged species will not be monophyletic for most loci (Hudson and Coyne 2002; Ting *et al.* 2000; Machado and Hey 2003) and, as a consequence, multiple gene genealogies will produce discordant trees (Dopman *et al.* 2005).

Although it is difficult to find monophyletic genes in recently diverged or incipient species, these species pairs offer the best opportunity for the study of speciation,

because post-speciation mutations, which are simply "noise" in the system, have not yet had time to accumulate. Furthermore recently diverged species are often capable of hybridization, and natural hybrid zones are windows on the evolutionary process (Harrison 1990). Hybrid zones can result in the introgression of one species alleles into another species background, leading to shared alleles and further erasing monophyletic relationships. Although alleles at many, even most, loci are free to introgress, some will be incompatible with the other species' genetic background causing hybrid unfitness. These so-called speciation or barrier genes will be prevented from crossing species boundaries, and exclusivity will be maintained at these loci. This will result in a semi-permeable species barrier where permeability will vary from locus to locus (Barton and Hewitt 1981; Harrison 1990; Wu 2001).

Barrier genes will play a crucial role in driving incipient species into fully differentiated genetic entities (Wu and Ting 2004), but they may not be easy to locate and study. A class of genes that is particularly promising are genes that encode accessory gland proteins. In animals with internal fertilization, these proteins are transferred to females along with sperm and play a very important role in reproductive interactions and possibly in the evolution of reproductive isolation. Accessory gland proteins have been shown to induce peristalsis of the female reproductive tract (Davey 1958), are essential for storage and capacitation of sperm (Neubam and Wolfner 1999; Bloch Qazi and Wolfner 2003) and can manipulate female's reproductive behavior to improve the male's fertilization success and hence may be agents of sexual conflict (Lung *et al.* 2002, Wolfner 2002, Chapman *et al.* 2001). Furthermore many of these proteins have been shown to be under positive selection and to accumulate genetic differences between populations at a high rate (Pahuis *et al.* 2003; Swanson *et al.* 2001; Civetta and Singh 1995).

Studying accessory gland genes in recently diverged species that continue to hybridize offers a great opportunity to investigate the origins of barriers to gene exchange. To investigate the evolution of barriers to gene exchange, I studied closely related hybridizing field crickets, *Gryllus firmus* and *G. pennsylvanicus*, using behavioral studies (Chapter 2), *Wolbachia* curing (Chapter 3) and gene genealogies of accessory gland genes (Chapter 4).

The Gryllus *hybrid zone*

The North American field crickets *Gryllus firmus* and *G. pennsylvanicus* are very recently diverged. Mitochondrial DNA sequence divergence is less than 1% (Willet *et al.* 1997) and neither species is an exclusive group (i.e. monophyletic group) based on four nuclear gene introns (Broughton and Harrison 2003). Genealogical patterns and shared ancestral polymorphisms among three species, *G. firmus, G. pennsylvanicus* and *G. ovisopis*, suggests that speciation events occurred recently and rapidly relative to effective population sizes (Broughton and Harrison 2003). These findings suggest that speciation can occur even if only a small fraction of the genome becomes differentiated. Genes associated with reproduction such as accessory gland protein genes, are good candidates to explain the evolution of reproductive isolation and speciation.

The two field crickets, *G. firmus* and *G. pennsylvanicus*, form an extensive well-characterized hybrid zone (Harrison and Arnold 1982; Harrison 1983; Harrison 1985). Postmating barriers include a one-way incompatibility between the two species. No offspring are produced from crosses of *G. firmus* females and *G. pennsylvanicus* males, but the reciprocal cross produces viable and fertile offspring (Harrison 1983). There is evidence that premating barriers are also important because F1 hybrids are rare in the hybrid zone, where most individuals resemble parental types and females prefer to mate with conspecific males (Ross and Harrison 2002; Harrison and Bogdanowicz 1997;

Maroja *et al.* chapter 2). Furthermore most of the offspring from hybrid zone field-inseminated *G. pennsylvanicus* females as well as from laboratory crosses in which *G. pennsylvanicus* females are paired with conspecific and heterospecific males, are non-hybrid (Harrison 1986; Harrison and Rand 1989).

Since there is no evidence of sperm precedence (assortative fertilization) between these species (G. Hume personal communication), this pattern is probably due to positive assortative mating. Additionally *G. firmus* and *G. pennsylvanicus* are partially isolated due to habitat preferences; the former occurs in sandy soils and the latter in loam soils (Ross and Harrison 2002). The existence of many barriers impeding gene flow, including the unidirectional post-mating incompatibility, makes this an excellent system for the study of speciation.

Is Wolbachia *causing reproductive incompatibility?*

It has been suggested that the one-way incompatibility observed between *G. firmus* and *G. pennsylvanicus* is caused by *Wolbachia* (Giordano *et al.* 1997). *Wolbachia* are a group of intracellular alpha proteobacteria parasites that infect many groups of arthropods (Werren *et al.* 1995). These bacteria are associated with reproductive incompatibilities (cytoplasmic incompatibilities - CI) in a wide range of insect species and populations (Breeuwer *et al.* 1992; O'Neil *et al.* 1992). When an infected male mates with an uninfected female no offspring is produced but uninfected males are fertile with both infected and uninfected females. Thus, the one-way incompatibility observed between the hybridizing field-crickets could be caused by *Wolbachia* if *G. pennsylvanicus* was infected and *G. firmus* was uninfected or was infected with a different strain that did not cause reproductive incompatibility.

Although this explanation sounds compelling, the patterns of infection observed in *G. firmus* and *G. pennsylvanicus* appear to be inconsistent with *Wolbachia*-induced

cytoplasmic incompatibility (Mandel *et al.* 2001). Mandel *et al.* (2001) showed that *Gryllus pennsylvanicus* is infected with one or two strains of *Wolbachia* (*w*G1 and/or *w*G2) while *G. firmus* is infected with *w*G1. Since many *G. pennsylvanicus* are singly infected with *w*G1 (also found in *G. firmus*) some of the crosses between female *G. firmus* and male *G. pennsylvanicus* should be fertile. However no offspring was ever produced in hundreds of laboratory crosses. Furthermore it has been shown that *G. pennsylvanicus* males are not able to induce oviposition in *G. firmus* to the same extent as conspecific males (see Chapter 2). This result is not expected for *Wolbachia*-induced cytoplasmic incompatibility since its effects are post-zygotic (Werren *et al.* 1995). In chapter 3 I show that infection patterns are even more complex with *w*G1 and *w*G2 present in both species. Furthermore, cured (uninfected) male *G. pennsylvanicus* are still unable to produce hybrid offspring. My results show that *Wolbachia* is not the cause of the one-way reproductive incompatibility.

Structure of this dissertation

This dissertation is composed of four chapters, including this introduction. Three subsequent chapters are written in scientific paper format and are relatively independent from each other.

In Chapter 2 I examine behavioral barriers to gene exchange between *G. firmus* and *G. pennsylvanicus* and fertility costs to *G. firmus* females from mating with both heterospecific and conspecific males (because of the one-way reproductive incompatibility, all offspring produced from *G. firmus* are sired by conspecific males). I document strong behavioral premating barriers in both species; there is a dramatic difference in time to mate with conspecific and heterospecific males and a higher failure to mate with heterospecific males. In spite of the preference to conspecific males, there are no apparent fecundity or fertility costs for *G. firmus* females mated with both

conspecific and heterospecific males. I also document a failure of heterospecific males to induce normal oviposition in *G. firmus* females. This failure could be due to rapid evolution of accessory gland proteins in *G. firmus* and/or *G. pennsylvanicus* and may serve as a barrier to gene exchange.

In Chapter 3 I examine whether *Wolbachia* causes cytoplasmic incompatibility. I use a combination of curing experiments and intra- and interspecific crosses, microscopy analysis to examine presence/absence of *Wolbachia* in reproductive tissues, and quantification of *Wolbachia* loads in *G. firmus* and *G. pennsylvanicus* using Real-Time PCR. I conclude that *Wolbachia* infections are not the cause of the one-way reproductive incompatibility between *G. firmus* and *G. pennsylvanicus*

In Chapter 4 I show discordant genealogical patterns and differential introgression rates across the genome and discuss two potential barrier genes. These two candidate barrier genes are both accessory gland expressed genes which encode products that proteomic analysis has shown to be present in spermatophores. They show near zero introgression between the two species and have more structured species trees; they are also the only two seminal protein loci under selection. I also used the genealogical data to examine the demographic history of the field crickets and the current structure of the hybrid zone.

REFERENCES

Bailey, R.I., C.D. Thomas, and R.K. Butlin. 2004. Premating barriers to gene exchange and their implications for the structure of a mosaic zone between *Chorthippus brunneus* and *C. jacobsi* (Orthoptera: Acrididae). J. Evol. Biol. 17: 108-119.

Barbash, D., D. Sinno, A. Tarone, and J. Roote. 2003. A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. Proc. Natl. Acad. Sci. USA 100: 5302-5307.

Barton, N.H., and G.M. Hewitt. 1981. Hybrid zones and speciation. Pp 109-145 *in* Evolution and Speciation (Atchley, W.R., and D.S. Woodruf, Eds). Cambridge University Press, Cambridge, UK.

Bert, T.M., D.M. Hesselman, W.S. Arnold, W.S. Moore, H. Cruz-Lopez, and D.C. Marelli. 1993. High frequency of gonadal neoplasia in a hard clam (Mercenaria spp) hybrid zone. Mar.Biol. 117: 97-104.

Bloch Qazi, M.C.B., and M.F. Wolfner. 2003. An early role for the *Drosophila melanogaster* male seminal protein Acp36DE in female sperm storage. J. Exp. Biol. 206: 3521-3528.

Breeuwer, J.A.J., R. Stouthamer, D.A. Burns, D.A. Pelletier, W.G. Weisburg and J.H. Werren. 1992. Phylogeny of cytoplasmic incompatibility microorganisms in the parasitoid wasp genus *Nasonia* (Hymenoptera: Pteromalidae) based on 16s ribosomal DNA sequences. Insect Molec. Biol. 1: 25-36.

Brideau, N.J., H.A. Flores, J. Wang, S. Maheshware, X. Wang, and D.A. Barbash. 2006. Two Dobzjansky-Muller genes interact to cause hybrid lethality in *Drosophila*. Science 314: 1292- 1295.

Broughton, R.E., and R.G. Harrison. 2003. Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. Genetics 163: 1389-1401.

Burke, J.M., S.E. Carney, and M.L. Arnold. 1998. Hybrid fitness in Louisiana irises: Analysis of parental and F1 performance. Evolution 52: 37-43.

Burke, J.M., and M.L. Arnold. 2001. Genetics and the fitness of hybrids. Annu. Rev. Genet. 35: 31-52

Chapman, T., L.A. Herndon, Y. Heifetz, L. Partridge, and M.F. Wolfner. 2001. The Acp26Aa seminal fluid protein is a modulator of early egg-hatchability in *Drosophila melanogaster*. Proc. Roy. Soc. Lond. B 268: 1647-1654.

Civetta A., and R.S. Singh. 1995. High divergence in reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *D. virilis* group species. J. Mol. Evol. 41:1085-1095.

Coyne, J.A., and Orr, H.A. 2004. *Speciation*. Sunderland, MA: Sinauer Associates, Inc.

Cruzan M.B., M.L. Arnold. 1994. Assortative mating and natural selection in an *Iris* hybrid zone. Evolution 48: 1946-1958.

Davey, D.G. 1958. The migration of spermatozoa in the female of *Rhodnius prolixus*. J. Exp. Biol. 35:694-701.

Dobzhansky, T. 1937. *Genetics and the Origin of Species*. Columbia Univ. Press, New York.

Dopman, E.B., L. Peréz, S.M. Bogdanowicz, and R.G. Harrison. 2005 Consequences of reproductive barriers for genealogical discordance in the European corn borer. Proc. Natl. Acad. Sci. USA. 102: 14706-14711.

Giordano, R., J.J. Jackson and H.M. Robertson. 1997. The role of *Wolbachia* bacteria in reproductive incompatibilities and hybrid zones of *Diabrotica* beetles and *Gryllus* crickets. Proc. Natl. Acad. Sci. USA. 94: 11439-11444.

Grahame, J.W., C.S. Wilding and R.K. Butlin. 2006. Adaptation to a steep environmental gradient and an associated barrier to gene exchange in *Littorina saxatilis.* Evolution 60: 268-278.

Harrison, R.G. 1983. Barriers to gene exchange between closely related cricket species. I. Laboratory hybridization studies. Evolution 37: 245-251.

Harrison, R.G. 1985. Barriers to gene exchange between closely related cricket species. II. Life cycle variation and temporal isolation. Evolution 39: 244-259.

Harrison, R.G. 1986. Pattern and process in a narrow hybrid zone. Heredity 56: 337-349.

Harrison, R.G., and D.M. Rand. 1989. Mosaic hybrid zones and the nature of species boundaries. Pp 111-133 *in* Speciation and its consequences (Otte, D., and J.A. Endler, Eds.). Sinauer, Sunderland, MA.

Harrison, R.G., and J. Arnold 1982. A narrow hybrid zone between closely related cricket species. Evolution 36: 535-552.

Harrison, R.G., and S.M. Bogdanowicz. 1997. Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. Evolution 51: 493-505.

Harrison, R.G. 1990. Hybrid zones: windows on evolutionary processes. Pp 69-128 *in* Oxford Surveys in Evolutionary Biology Vol 7 (Futuyma D., and J. Antonovics, Eds). Oxford University Press, New York.

Howard, D.J., P.G. Gregory, J. Chu, and M.L. Cain. 1998. Conspecific sperm precedence is an effective barrier to hybridization between closely related species. Evolution 52: 511-516.

Hudson, R.R and J.A. Coyne. 2002. Mathematical consequences of the genealogical species concept. Evolution 56: 1557-1565.

Hudson, R.R. 1990. Gene genealogies and the coalescent process. Pp 1-44 *in* Oxford Surveys in Evolutionary Biology Vol 7 (Futuyma, D., and J. Antonovics, Eds.). Oxford University Press, New York.

Hudson, R.R. 1992. Gene trees, species trees and the segregation of ancestral alleles. Genetics 131: 509-512.

Kliman, R.M., P. Andolfatto, J.A. Coyne, F. Depaulis, M. Kreitman, A.J. Berry, J. McCarter; J. Wakeley and J. Hey. 2000. The population genetic of the origin and diversification of the *Drosophila simulans* complex species. Genetics 156: 1913-1931.

Lung, O., U. Tram, C. Finnerty, M. Eipper-Mains, J.M. Kalb, and M.F. Wolfner. 2002. The *Drosophila melanogaster* seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression. Genetics 160: 211-224.

MacCallum, C.J., B. Nurnberger, N.H. Barton, and J.M. Szymura. 1998. Habitat preference in *Bombina* hybrid zone in Croatia. Evolution 52: 227-239.

Machado, C.A., and J. Hey. 2003 The causes of phylogenetic conflict in a classic *Drosophila* species group. Proc. R. Soc. Lond. B 270: 1193-1202.

Mandel, M.J., C.L. Ross, and R.G. Harrison. 2001. Do *Wolbachia* infections play a role in unidirectional incompatibilities in a field cricket hybrid zone? Mol. Ecol. 10: 703-709.

Masly J.P., C.D. Jones, M.A.F. Noor, and H.A. Orr. 2006. Gene transposition as a cause of hybrid sterility. Science 313: 1448-1450.

Neigel, J.E., and J.C. Avise. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. Pp 515-534 *in* Evolutionary Processes and Theory (Karlin S., and E. Nevo, Eds). Academic Press, New York.

Neubaum, D.M., and M.F. Wolfner. 1999. Mated *Drosphila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. Genetics 153:845-857.

Nichols, R. 2001. Gene trees and species trees are not the same. TREE 16: 358-364.

O'Neill, S.L., R. Giordano, A.M.E. Colbert, T.L. Karr, and H.M. Robertson. 1992. 16s rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. Proc. Natl. Acad. Sci. USA. 89: 2699-2702.

Orr, H.A., J.P. Masly, and N.P. Phadnis. 2007. Speciation in *Drosophila*: from phenotypes to molecules. J. Hered 98: 103-110.

Panhuis, T.M., W.J. Swanson, and L. Nunney. 2003. Population Genetic of accessory gland proteins and sexual behavior in *Drosophila melanogaster* populations from Evolution canyon. Evolution 57: 2785-2791.

Presgraves, D., L. Balagopalan, S. Abmayr, and H. Orr. 2003. Adaptive evolution drives divergence of hybrid incompatibility gene between two species of *Drosophila*. Nature 243: 715-719.

Rieseberg, L.H., J. Whitton, and K. Gardner. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics 152: 713-727.

Ross, C.L., and R. G. Harrison. 2002. A fine-scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus*. Evolution 56: 2296-2312.

Saetre, G.P., M. Kral, and S. Bures. 1997. Differential species recognition abilities of males and females in a flycatcher hybrid zone. J. Avian Biol. 28: 259-263.

Swanson, W.J., A.G. Clark, H.M Waldrip-Dail, M.F. Wolfner, and C.F. Aquadro. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc Natl Acad Sci USA. 98:7375-7379.

Ting, C., S.-C. Tsaur, and C.-I. Wu. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. Science 282: 1501-1504.

Ting, C.-T., S.-C. Tsaur, and C.-I. Wu. 2000. The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus.* Proc. Natl. Acad. Sci. USA. 97: 5313-5316.

Wang, R.-L., J. Wakeley, and J. Hey. 1997. Gene flow and natural selection in the origin of *Drosophila pseudobscura* and close relatives. Genetics 147: 1091-1106.

Werren, J.H., W. Zhang, and L.R. Guo. 1995. Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. Proc. R. Soc. Lond. B. 261: 55-71.

Willet, C., M.J. Ford, and R.G. Harrison. 1997. Inferences about the origin of a field cricket hybrid zone from a mitochondrial DNA phylogeny. Heredity 79: 484-494.

Wittbrodt J., D. Adam, B. Malitscheck, W. Maueler, F. Raulf, A. Telling, S.M. Robertson, and M. Schartl. 1989. Novel putative receptor tyrosine kinase encoded by melanoma-inducing *Tu* locus in *Xiphophorus.* Nature 341: 415-421.

Wolfner, M.F. 2002. The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in *Drosophila*. Heredity 88:85-93.

Wu, C.-I. 2001. The genic view of the process of speciation. J. Evol. Biol. 14: 851-865.

Wu, C.-I., and C.Tl Ting. 2004. Genes and speciation. Nat. Rev. Genet. 5: 114-122.

CHAPTER 2

MULTIPLE BARRIERS TO GENE EXCHANGE IN THE *GRYLLUS FIRMUS* - *GRYLLUS PENNSYLVANICUS* HYBRID ZONE.

ABSTRACT

Data on patterns of variation within hybrid zones, combined with studies of life history, mate choice, and hybrid performance, allow estimates of the importance of prezygotic and postzygotic barriers. We examine the role of behavioral barriers to gene exchange in the maintenance of a hybrid zone between North American field crickets *Gryllus firmus* and *G. pennsylvanicus*. We consider these barriers in the context of previous studies that documented temporal and ecological isolation and a one-way post-mating incompatibility. Our results reveal strong behavioral premating barriers, but no apparent fecundity or fertility costs for *G. firmus* females when they mate with both conspecific and heterospecific males. We also document a failure of heterospecific males to induce normal oviposition in *G. firmus* females. This failure could be due to rapid evolution of accessory gland proteins in *G. firmus* and/or *G. pennsylvanicus* and may serve as a barrier to gene exchange.

INTRODUCTION

Hybrid zones provide valuable insights into the operation and evolution of barriers to gene exchange between closely related species (Hewitt 1988; Harrison 1990). Many hybrid zones represent secondary contact between populations or species that have diverged in allopatry, and differences between hybridizing taxa will only persist if one or more barriers are strong enough to counteract the homogenizing effects of gene flow. Data on patterns of variation within hybrid zones, combined with studies of life history, mate choice, and hybrid performance in the laboratory,

allow estimates of the contribution of both prezygotic and postzygotic barriers.

Prezygotic barriers to gene exchange include ecological, temporal, behavioral and mechanical barriers that "prevent interspecific crosses" (Mayr 1963). Prezygotic barriers also include conspecific sperm precedence and cryptic female choice, in which sperm (pollen) transfer occurs but heterospecific sperm (pollen) cannot effect fertilization (Mayr 1963; Coyne and Orr 2004). Postzygotic barriers include hybrid inviability, infertility and F2 breakdown (Dobzhansky 1937; Mayr 1963), as well as ecological inviability and behavioral sterility (Coyne and Orr 2004).

In many taxa multiple reproductive barriers contribute to isolation (Coyne 1992; Schluter 2001; Price and Bouvier 2002; Ramsey *et al.* 2003), but the relative contribution of each barrier, as well as their importance in the speciation process often remain unknown. Only in a few model systems (e.g., sympatric species of *Mimulus* in North America; Ramsey *et al.* 2003) have quantitative estimates of individual "barrier strengths" been made.

Despite the obvious importance of and emphasis on postzygotic barriers (e.g, the extensive literature on Dobzhansky-Muller incompatibilities and Haldane's Rule; Coyne and Orr 2004), cases in which species differences persist only because of postzygotic barriers are thought to be relatively rare (Kirkpatrick and Ravigne 2002). Furthermore, premating barriers act early in the life cycle, and thus have the potential to reduce gene flow proportionally more than barriers that act later (Coyne and Orr 2004). Within hybrid zones, pre- and postzygotic barriers are equally important in unimodal zones (those in which hybrid genotypes are common), whereas bimodal hybrid zones (those in which parental types predominate) "are invariably coupled with strong assortative mating or assortative fertilization" (Jiggins and Mallet 2000). Thus, prezygotic barriers (and especially behavioral barriers) appear to be responsible for the deficiency of heterozygotes or "intermediate" individuals and the strong linkage

disequilibrium characteristic of bimodal hybrid zones (Harrison and Bogdanowicz 1997; Jiggins and Mallet 2000; Vines *et al.* 2003; Ross and Harrison 2006).

Within hybrid zones, the evolution of behavioral prezygotic barriers will depend on the cost to females of mating with heterospecific males. In the presence of postzygotic barriers, less fit or inviable hybrids are produced, and costs of heterospecific matings are high, potentially leading to reproductive character displacement (Dobzhansky 1940; Howard 1993; Coyne and Orr 2004). However, when females are polyandrous and there is strong sperm precedence and/or gametic incompatibilities, few or no hybrid offspring may be produced, substantially reducing the cost of mating with heterospecific males. Females who mate multiply are more likely to receive sperm from at least one conspecific male and thus ensure fertilization. In these cases the mating cost per se might be very low, especially if females gain direct benefits (e.g. access to resources; Andersson 1994) from mating with heterospecific males.

Here we examine the role of behavioral barriers to gene exchange in the maintenance of a bimodal mosaic hybrid zone between the North American field crickets *Gryllus firmus* and *G. pennsylvanicus*. We consider these barriers in the context of previous studies that have documented temporal and ecological isolation and an apparent one-way gametic incompatibility (Harrison 1983, 1985; Harrison and Rand 1989; Ross and Harrison 2002, 2006). We estimate the potential cost for *G. firmus* females of mating with heterospecific males. Our results reveal strong behavioral premating barriers, but no apparent cost for *G. firmus* females when they mate with both conspecific and heterospecific males.

The study system

The reproductive biology of field crickets has been very well studied. In many species females locate suitable mates based on variation in calling song (Hedrick 1986;

Simmons 1988; Wagner 1996). However, when females approach, males switch from calling song to courtship song and initiate production of a number of chemical signals (Loher and Dambach 1989). In gryllid crickets, forced copulation is impossible because the female must mount the male; both sexes cooperate in the transfer of the spermatophore. Females are polyandrous (Solymar and Cade 1990; Bretman and Tregenza 2005) and are able to store sperm from many mates in a single elastic spermatheca (Simmons 1986; Bretman and Tregenza 2005). Unlike many other insects, field cricket females appear to benefit from multiple mating through both direct (increased lifetime fecundity) and indirect (i.e., genetic) benefits (Simmons 1988; Burpee and Sakaluk 1993; Wagner *et al.* 2001; Sakaluk *et al.* 2002; Ivy and Sakaluk 2005).

Here we focus on the field crickets *Gryllus firmus* and *G. pennsylvanicus,* which form an extensive hybrid zone (Harrison and Arnold 1982; Harrison and Bogdanowicz 1997) in which there is a one-way incompatibility that restricts gene flow between the species (Harrison 1983). No offspring are produced from crosses of *G. firmus* females and *G. pennsylvanicus* males, but the reciprocal cross produces viable and fertile offspring (Harrison 1983). Therefore, the potential importance and consequences of behavioral barriers are different for the two reciprocal crosses. There is evidence that prezygotic barriers are also important in hybrid zone populations, because F1 hybrids are rare and most individuals from the hybrid zone resemble parental types (Harrison and Bogdanowicz 1997; Ross and Harrison 2002). Thus, the cricket hybrid zone is clearly bimodal. Habitat associations reduce encounter rates (and therefore gene flow) between *G. firmus* and *G. pennsylvanicus*; the former occurs on sandy soils and the latter on loam soils (Rand and Harrison 1989; Ross and Harrison 2002, 2006). However, other barriers must also be present, because in the laboratory *G. pennsylvanicus* females housed with males of both species produce offspring sired primarily by conspecific males (Harrison and Rand 1989). These data suggest positive assortative mating since there is no evidence

of conspecific sperm precedence (assortative fertilization) in these species (G. Hume personal communication). Despite evidence pointing to the existence of behavioral barriers to gene exchange, no experiments on female mate choice have been performed. To fill this gap, we observed single pair crosses and recorded time to mating and rejection rate of males for *G. firmus* and *G. pennsylvanicus* females presented with both conspecific and heterospecific males.

We also used measures of fecundity and fertility to investigate the costs to *G. firmus* females of mating with heterospecific males. *Gryllus firmus* females do not produce hybrid offspring when mated only to heterospecific males (Harrison 1983), but it is unclear whether there are costs when a female mates with both conspecific and heterospecific males. Finally, we examined the fecundity of *G. firmus* females mated only to heterospecific males. Anecdotal evidence suggests that *G. firmus* females mated only to heterospecific males deposit very few or no eggs, perhaps because heterospecific males do not trigger normal oviposition behavior.

MATERIALS AND METHODS

We collected late instar *Gryllus firmus* nymphs in Guilford, CT (41°15'; -72°42') and *G. pennsylvanicus* nymphs in Ithaca, NY (42°24'; -76°31'). Both species were collected during August-September, 2003-04. We sorted the crickets by sex and species and maintained them in plastic cages (30×16×9 cm) with food (Purina Cat Chow®), a water vial, and cardboard for shelter. The cages were kept at 25°C, 12: 12 light: dark.

Mating trials with Gryllus firmus females

Seven to eight days old adult *G. firmus* virgin females were randomly assigned to one of the six following treatments (Figure 2.1). In treatment F, females ($n_F = 15$) were not given access to males. In treatments FF and FFF females either mated once with a

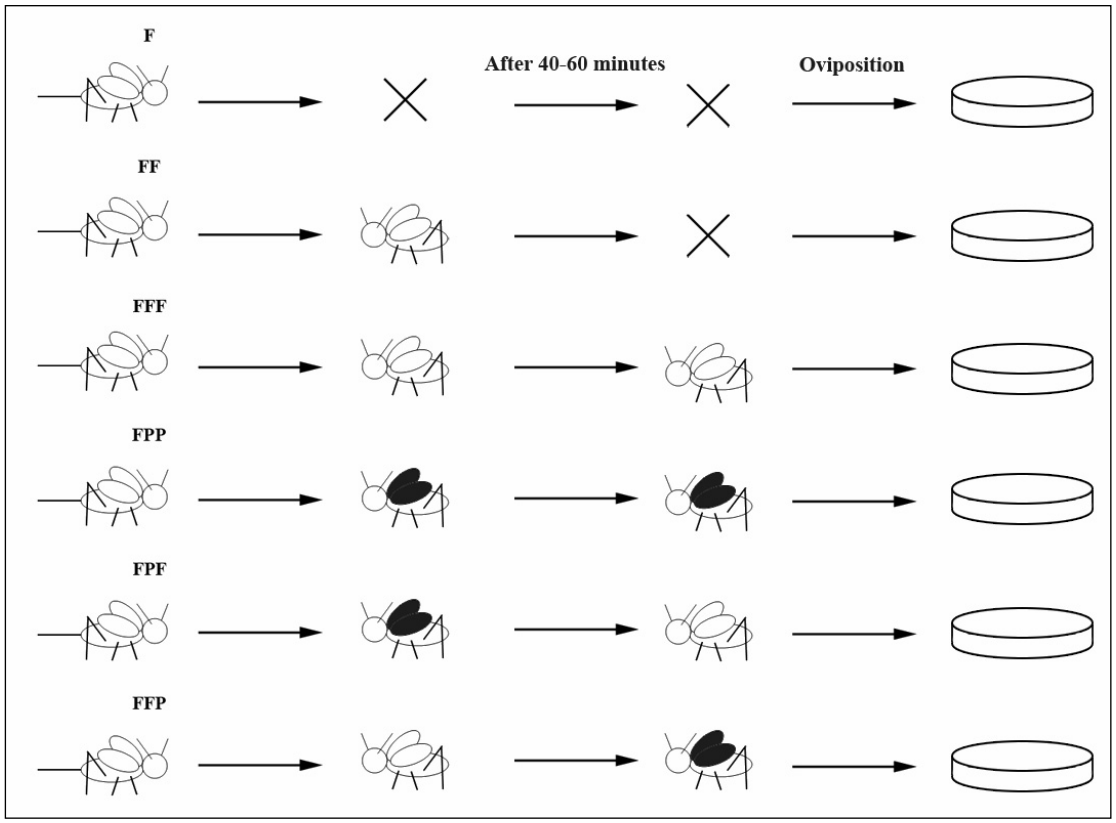


Figure 2.1 Experimental protocol for *Gryllus firmus* females (similar protocol for *G. pennsylvanicus* females). Crickets in the left column are females (with ovipositor). Crickets with black wings represent heterospecific males. See text for details.

conspecific male (n_{FF} = 15) or twice with two different conspecific males (n_{FFF} = 15). In treatment FPP females mated twice with two different heterospecific males (n_{FPP} = 12). In treatment FPF females mated first with a heterospecific male (*G. pennsylvanicus*) and second with a conspecific one (n_{FPF} =15). Finally, in treatment FFP (n_{FFP} = 12) each female was mated first to a conspecific male and second to a heterospecific one. All matings started at 7:30am. All the males used in the mating trials had been adult for 7-12 days. Males were chosen at random and used only once to avoid pseudo-replication. All individuals were sized by measuring pronotal width to the nearest 0.1 mm using the same pair of vernier calipers.

To initiate the mating trials, each virgin female was placed with a first male in a mating chamber consisting of a 10 cm Petri dish containing moist filter paper. If no mating (i.e. spermatophore transfer) occurred during the first hour, this first male was removed and a second male of the same species was introduced into the mating chamber for no more than one hour. If the female remained unmated after this time, we scored the mating trial as failed and excluded both males and female from subsequent trials. Single-mated FF females were isolated individually in plastic oviposition chambers (30×15×8 cm), provided with food, water and a Petri dish of sterilized soil as oviposition substrate. With other single-mated females we proceeded with the second part of the mating trials as follows: immediately after the female detached the spermatophore from her first mating (about 40 minutes after mating), she was transferred to a new mating chamber and placed with a second male for an hour. As in the first part of the mating trials, if mating did not take place during this time, females were exposed to a new second male of the same species for another hour. All the focal females that did not mate during the second part of the trial were removed from the analyses. After mating for a second time, females were individually isolated as described above. Food and water in the individual oviposition chambers were replaced twice a week and mortality was scored every other day.

Oviposition dishes were incubated for a maximum of 40 days at 25°C and then placed in a refrigerator at 4°C for 102 days to insure synchronous hatch of nymphs (Harrison, 1985). Lifetime fecundity was assessed by counting all of the eggs laid by each female. Eggs were first separated from the oviposition substrate using a series of sieves, and then counted under a stereoscopic microscope. To estimate fertility, samples of 100 eggs were taken from each female. For those females that laid less than 100 eggs ($n_F = 13$; $n_{FF} = 1$; $n_{FFF} = 0$; $n_{FPF} = 1$; $n_{FFP} = 2$; $n_{FPP} = 6$) fertility was assessed using the entire clutch. Fertility was estimated as the proportion of eggs that successfully hatched. After a diapause period of 102 days at 4°C, all eggs were incubated at room temperature. Hatching began 11 days after the eggs were removed from the refrigerator. Two weeks after the first nymphs hatched, the number of offspring was determined. No eggs hatched after this period.

Mating trials with Gryllus pennsylvanicus females

Seven to eight days old adult *G. pennsylvanicus* virgin females were randomly assigned to one of the four different treatments. In treatment PPP, females were sequentially mated to two different conspecific males ($n_{PPP} = 9$). In treatment PFF, females mated sequentially with two different heterospecific males ($n_{PFF} = 9$). In treatment PPF, females were first mated to a conspecific male and then to a heterospecific one ($n_{FPF} = 15$). Finally, in the last treatment PFP, females mated first with a heterospecific male and subsequently with a conspecific one ($n_{FFP} = 12$). The mating protocol was the same as the one described above for the *G. firmus* mating trials. For this experiment both males and females were eliminated after mating (no eggs collected).

Data analyses

Mating trial data sets were analyzed using generalized linear models (GLMs). Initial inspection of the time to mate data showed a constant coefficient of variation.

Therefore we fitted our data to GLMs with Gamma errors and a log link function (Crawley 1993).

Data on fertility represent proportions and therefore normal linear models are potentially inappropriate (Lindsey 1995). For this reason we first fitted our data to GLMs with binomial errors using a logit link function. However a visual inspection of the error structure of this model revealed it to be inappropriate for our data set. Therefore, we fitted a model with a Poisson error structure and log link (Crawley 1993). In those cases in which our sampling variance exceeded the theoretical (i.e. model based) variance (McCullagh and Nelder 1989; Burnham and Anderson 1998) we fitted our data to a negative binomial model, which includes a random term reflecting unexplained between-subject differences (Gardner *et al.* 1995).

Residuals of all the performed GLMs were analyzed by visual inspection and no deviations from normality were observed. No significant outliers were found using Cook's statistics values. The effects of all dependent variables on the response variables were tested using log-likelihood ratio tests comparing the deviance of a model including and excluding the factor being tested. Analyses were performed with R. 1.9.0 (R Development Core Team, 2004). All results are shown as (mean ± SE).

RESULTS

Time to mate and mating trial failures

For *Gryllus firmus* females, time to mate with heterospecific males (42.9 ± 4.4 min) was significantly longer than time to mate with conspecific males (7.4 ± 1.0 min; Figure 2.2). In matings involving virgin females (mating for the first time), time to mate is explained only by the first male species (test of full model; F_{ratio}= 67.38, $P < 0.0001$; Table 2.1). In matings involving previously mated females, time to mate is explained by

Table 2.1 Results of generalized full linear model for time to mate of *G. firmus* virgin females. Only double mating treatments are included (FFF, FPF, FFP, and FPP). Significant terms are shown in bold.

	deviance	d.f.	LLR*	P
Time to mate for virgin females	71.01	55	—	—
Male species	—	1	67.21	6.19e-11
Male size	—	1	0.13	0.71
Male species × male size	—	1	0.01	0.94

*LLR = value of the log likelihood ratio test (F test)

Table 2.2 Results of generalized full linear model for time to mate for previously mated *G. firmus* females. Only double mating treatments are included (FFF, FPF, FFP, and FPP). Significant terms are shown in bold.

	deviance	d.f.	LLR*	P
Time to mate with second male	77.17	54	—	—
1st male species	—	1	5.33	0.03
2nd male species	—	1	69.24	3.55e-10
1st male size	—	1	0.01	0.90
2nd male size	—	1	18.97	9.34e-05
1st male species× 2nd male species	—	1	0.27	0.60
1st male species× 1st male size	—	1	2.24	0.14
2nd male species × 1st male size	—	1	0.69	0.41
1st male species × 2nd male size	—	1	0.09	0.76
2nd male × 2nd male size	—	1	3.73	0.06
1st male size × 2nd male size	—	1	2.45	0.12
1st male species×2nd male species×1st male size	—	1	0.61	0.44
1st male species×2nd male species×2nd male size	—	1	8.62	0.006
1st male species×1st male size×2nd male size	—	1	2.20	0.15
2nd male species×1st male size×2nd male size	—	1	0.22	0.64
1st male species×2nd male species×1st male size ×2nd male size	—	1	1.52	0.23

*LLR = value of the log likelihood ratio test (F test)

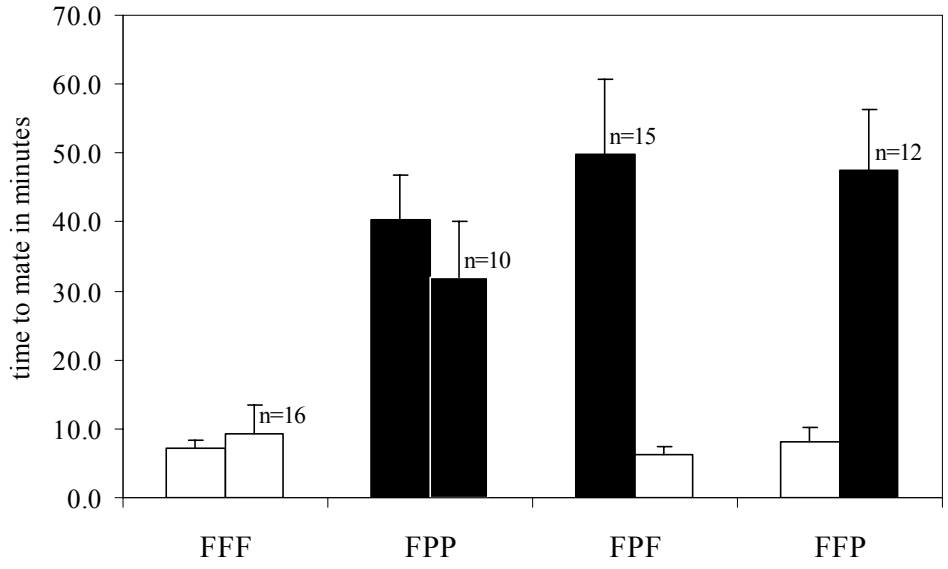


Figure 2.2 Mean and standard error of *G. firmus* female time to mate with first and second male for each double mating treatment (n=sample size). White bars represent conspecific males. Black bars represent heterospecific male.

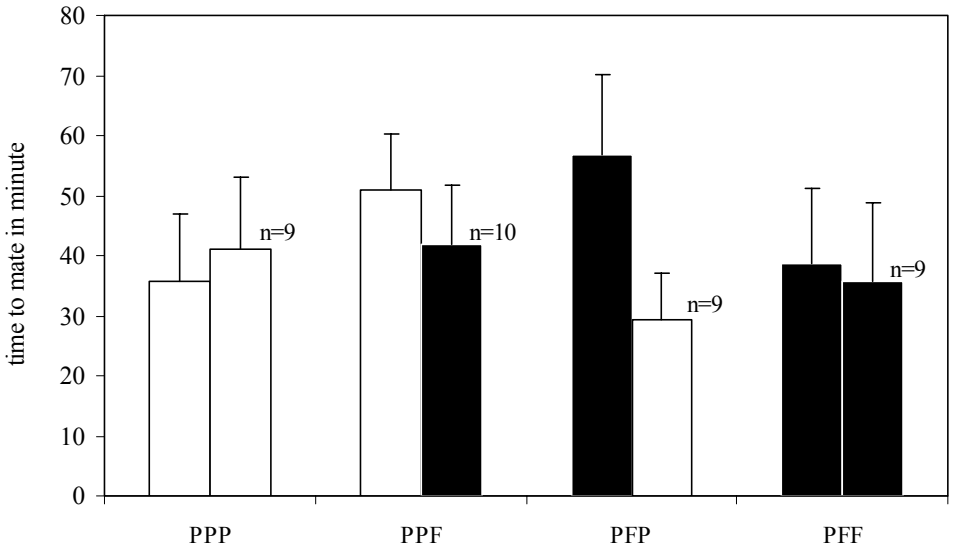


Figure 2.3 Mean and standard error of *G. pennsylvanicus* female time to mate with first and second male for each double mating treatment (n=sample size). White bars represent conspecific males. Black bars represent heterospecific male.

first and second male species as well as second male size. In addition these factors interacted significantly (test of full model; F_{ratio}= 119.19, $P < 0.0001$; Table 2.2). Time to remate was less if females first mated with a heterospecific male; surprisingly, females remated faster to smaller males. Furthermore, significantly more virgin *G. firmus* females failed to mate with heterospecific males (22.2%, n=8) than with conspecific males (n=0, Fisher's exact test $P = 0.001$).

In contrast, *G. pennsylvanicus* female time to mate was only slightly longer with heterospecific males (43.2 ± 4.8 min) than conspecific ones (39.6 ± 5.7 min; Figure 2.3), and none of the variables measured in this study seem to have had a significant effect on the time to mate; this is the case both for virgins and mated females (Table 2.3 and Table 2.4). Independent of the species of the male, virgin *G. pennsylvanicus* were more reluctant to mate (38.3% failures, n=41) than virgin *G. firmus* females (10.4% failures, n=8, Fisher exact test $P < 0.0001$). However, as with *G. firmus*, significantly more virgin females failed to mate heterospecific males (47.3%, n=26) than conspecific males (28.5%, n=15; Fisher's exact test P = 0.02).

Fertility and fecundity in G. firmus females

There was no difference in lifetime fecundity (Figure 2.4) or fertility (Figure 2.5) for *G. firmus* females mated to at least one conspecific male (FF, FFF, FPF and FFP). Lifetime fecundity (i.e. total number of eggs laid) was only significantly affected by female lifespan (test of full model; χ^2=34.94, 15 d.f., $P < 0.005$, Table 2.5). Fertility (number of hatchlings in a sample of 100 eggs) was not significantly affected by any factor measured in this study (Table 2.6).

The numbers of eggs deposited by *G. firmus* virgin females (F, 47.5 ± 31 eggs) and females mated only to heterospecific males (FPP, 181.7 ± 53.7 eggs) were much lower than numbers of eggs from females mated to at least one conspecific male (FF, FFF,

Table 2.3 Results of generalized full linear model for time to mate of *G. pennsylvanicus* virgin females. All double mating treatments are included (PFF, PPF, PFP, and PPP).

	deviance	d.f.	LLR*	P
Time to mate for virgin females	29.63	35	—	—
Male species	—	1	0.04	0.83
Male size	—	1	0.85	0.36
Male species × male size	—	1	4.04	0.053

*LLR = value of the log likelihood ratio test (F test)

Table 2.4 Results of generalized full linear model for time to mate for previously mated *G. pennsylvanicus* females. All double mating treatments are included (PFF, PPF, PFP, and PPP).

	deviance	d.f.	LLR*	P
Time to mate with second male	26.98	34	—	—
1st male species	—	1	2.08	0.16
2nd male species	—	1	0.11	0.73
1st male size	—	1	1.14	0.29
2nd male size	—	1	0.11	0.75
1st male species× 2nd male species	—	1	0.01	0.93
1st male species× 1st male size	—	1	1.79	0.19
2nd male species × 1st male size	—	1	0.11	0.75
1st male species × 2nd male size	—	1	2.22	0.15
2nd male species × 2nd male size	—	1	0.26	0.61
1st male size × 2nd male size	—	1	0.04	0.85
1st male species×2nd male species×1st male size	—	1	0.36	0.56
1st male species×2nd male species×2nd male size	—	1	1.08	0.31
1st male species×1st male size×2nd male size	—	1	0.38	0.54
2nd male species×1st male size×2nd male size	—	1	0.15	0.70
1st male species×2nd male species×1st male size ×2nd male size	—	1	1.42	0.25

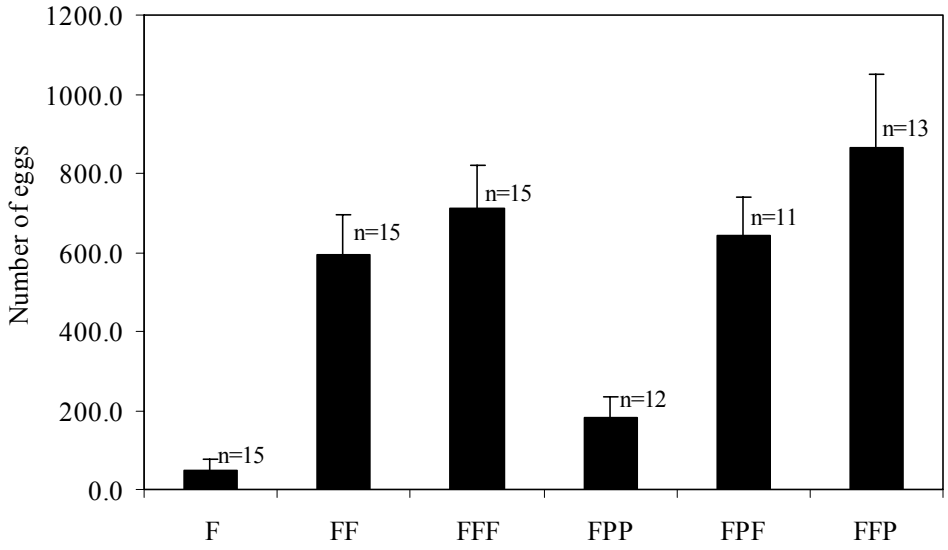


Figure 2.4 Mean and standard error of *G. firmus* fecundity for each treatment (n=sample size).

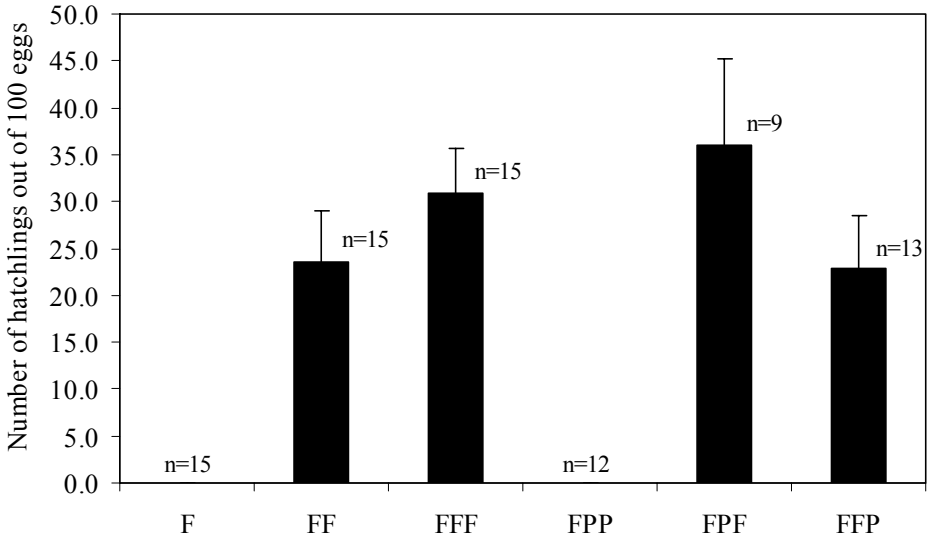


Figure 2.5 Mean and standard error of *G. firmus* fertility for each treatment (n=sample size).

Table 2.5 Results of generalized linear full model for fecundity of *G. firmus* females from treatments FF, FFF, FPF and FFP. Significant terms are shown in bold.

	Deviance	d.f.	LLR*	P
Fecundity	87.98	49	—	—
Treatment	—	3	5.31	0.15
Female lifespan	—	1	23.28	1.4e-6
Female size	—	1	0.49	0.49
Treatment × Female lifespan	—	3	1.47	0.69
Treatment × Female size	—	3	0.34	0.95
Female lifespan × Female size	—	1	0.04	0.85
Treatment × Female lifespan × Female size	—	3	4.02	0.26

*LLR = value of the log likelihood ratio test (χ^2 test)

Table 2.6 Results of generalized linear full model for fertility of *G. firmus* females from treatments FF, FFF, FPF and FFP.

	deviance	d.f.	LLR*	P
Fertility	68.48	47	—	—
Treatment	—	3	1.33	0.72
Female lifespan	—	1	0.34	0.56
Female size	—	1	0.39	0.53
Treatment × Female lifespan	—	3	1.32	0.72
Treatment × Female size	—	3	3.07	0.38
Female lifespan × Female size	—	1	1.56	0.21
Treatment × Female lifespan × Female size	—	3	0.42	0.94

*LLR = value of the log likelihood ratio test (F test)

FPF and FFP, 702.0 ± 61.7 eggs, Figure 2.2). Despite the low fecundity of the two treatments (F and FPP), there was a significant difference between them. Females mated to heterospecific males laid more eggs than virgin females; fecundity was affected both by treatment (F or FPP) and female size (test of full model, χ^2=23.79, 7 d.f. $P < 0.005$, Table 2.7). However neither virgins nor females mated only to heterospecific males produced any offspring (Figure 2.5).

DISCUSSION

Here we report previously uncharacterized behavioral premating barriers to gene exchange in the *Gryllus firmus* - *G. pennsylvanicus* hybrid zone. These two cricket species exhibit very little differentiation in morphology or DNA sequence and are assumed to have diverged very recently. Indeed, analyses of molecular markers (allozymes, mtDNA sequences, nuclear RFLPs, nuclear gene intron sequences) have uncovered very few diagnostic differences (Harrison and Arnold 1982; Harrison and Bogdanowicz 1997), and gene genealogies often reveal absence of exclusivity and haplotype sharing between the species (Willett *et al.* 1997; Broughton and Harrison 2003). In contrast to the similarities in morphology and gene sequences, the ecology, behavior, and development of the two cricket species have apparently diverged substantially. These differences, including differences in mating behavior reported here, result in multiple barriers to gene exchange that act throughout the life history of *G. firmus* and *G. pennsylvanicus* (see Table 2.8).

Many insect hybrid zones are reported to have multiple trait differences that restrict gene flow (e.g., Mendelson and Shaw 2002; Ross and Harrison 2002; Bailey *et al.* 2004). In the geographically extensive *G. firmus* - *G. pennsylvanicus* hybrid zone, some barriers operate throughout the zone, whereas others vary geographically (Table 2.8). For example, the one-way incompatibility between *G. firmus* females and *G. pennsylvanicus*

Table 2.7 Results of generalized full linear model for fecundity of *G. firmus* females from F and FPP treatments (virgins or females mated only to *G. pennsylvanicus* males).

	deviance	d.f.	LLR*	P
Fecundity for F and FPP	53.01	24	—	—
Treatment	—	1	17.70	2.58e-05
Female lifespan	—	1	4.65	0.03
Female size	—	1	0.38	0.54
Treatment × female lifespan	—	1	0.01	0.98
Treatment × female size	—	1	0.44	0.50
Female lifespan × female size	—	1	0.60	0.44
Treatment × female lifespan × female size	—	1	0.01	0.94

*LLR = value of the log likelihood ratio test (χ^2 test)

Table 2.8 List of known pre- and post-mating barriers to gene exchange between G. *firmus* and *G. pennsylvanicus*.

	Barrier	Likely Mechanism	References
Pre	Ecogeographic isolation	Association with different soils.	Rand and Harrison 1989; Ross and Harrison 2002, 2006
	Temporal isolation	Differences in time of adult appearance (due to differences in development times.).	Harrison 1985
	Acoustic isolation	Differences in calling song.	Alexander 1957; Doherty and Storz 1992
	Time to mate	Differences in time to mate with conspecific and heterospecifics.	This chapter
Post	One way incompatibility	Gametic incompatibility in the heterospecific cross between *G. firmus* female and *G. pensylvanicus* male.	Harrison 1983

males has been shown to be characteristic of crickets from both Connecticut and Virginia, whereas a clear soil association has only been documented in Connecticut (Rand and Harrison 1989; Ross and Harrison 2002); temporal isolation (due to differences in development time) is observed in Virginia but not in Connecticut (Harrison 1985). None of these barriers acting alone is complete, but together they appear to severely restrict gene exchange; very few F1 individuals are found in mixed populations and the hybrid zone is clearly bimodal.

The mating trials reported here suggest that the two cricket species (at least in Connecticut) differ in mating behavior when males and females are paired in no choice experiments. We used two measures of mate preference, time to mate and proportion of trials in which spermatophore transfer occurred. Using the latter criterion, females of both species "prefer" conspecific males. However, only *G. firmus* females showed significant differences in the time to mate; these females mated readily with conspecific males, but took far longer to mate with heterospecific males. *G. pennsylvanicus* females were generally more reluctant to mate; frequency of spermatophore transfer but not time to mate differed depending on the species of male with which they were paired.

The phenotypic differences responsible for the observed mate choice remain unclear. Acoustic signals often can be used to distinguish morphologically similar species of Orthoptera and have been shown to play a role in premating isolation and female choice (Wells and Henry 1998; Mendelson and Shaw 2002; Bridle *et al.* 2006). In some crickets, females clearly respond preferentially to conspecific male song (Mendelson and Shaw 2002; Holzer *et al.* 2003, Saldamando *et al.* 2005), and it has been argued that differences in calling song could be important in the maintenance of the *G. firmus* - *G. pennsylvanicus* hybrid zone (Doherty and Storz 1992). Although there are slight differences in the calling song of these species (Alexander 1957; Harrison and Rand 1989; Doherty and Storz 1992), the courtship song of North American *Gryllus* species does not vary (Alexander

1968). Because each male-female pair in our experiments was housed in a small confined space, in which females were exposed only to courtship song, the premating barriers reported here are not due to differences in calling song. Variation in chemical cues (e.g., cuticular hydrocarbons) has been shown to differentiate closely related species of other insects (Hardy and Shaw 1983; Howard and Blomquist 2005; Nagamoto *et al.* 2005; Mullen *et al.* 2007) and may play a role in sexual selection and speciation in *Gryllus*.

The laboratory, no-choice experiments reported here show that females of both species prefer to mate with conspecific males. However, such experiments obviously fail to mimic situations in natural populations, and it might be argued that the results have little bearing on what goes on within the hybrid zone. Indeed, given the local abundance of field crickets in hybrid zone populations (personal observation), females are rarely in no-choice situations. But in the presence of multiple males, female reluctance (or failure) to mate with heterospecifc males should serve as a substantial barrier and make heterospecific matings rare in the wild. Indeed the difficulty of finding F1 individuals in the hybrid zone (Harrison and Bogdanowicz 1997) confirms the importance of the premating barriers in preventing gene flow, given that there is no conspecific sperm precedence (G. Hume personal communication). Although habitat isolation no doubt serves to reduce encounter rates between the two crickets, adults of both species are found together within single populations in Connecticut (Harrison 1986; Harrison and Bogdanowicz 1997), and behavioral barriers provide the only explanation for the persistent bimodal nature of the hybrid zone.

Individuals from hybrid zones are expected to evolve stronger assortative mating if there are costs to mating with heterospecifics (Liou and Price 1994). Furthermore mosaic bimodal hybrid zones may facilitate reproductive character displacement by providing an initial level of assortative mating by habitat use (Jiggins and Mallet 2000). However, because non-random mating is costly (Anderson 1994), females are expected to

be choosy only if the cost of mating with the "wrong" male is high. Here we did not find any costs in fecundity or fertility for *G. firmus* females mated to both conspecific and heterospecific males (Figures 2.1 and 2.2). Because these females do not produce hybrid offspring, they do not suffer the associated costs and could even benefit from heterospecific matings if there is an advantage in multiple mating (Wagner *et al.* 2001). However, as is evident from the strong preference for conspecific males, *G. firmus* females are choosy despite the apparent absence of fecundity or fertility costs and the possible benefits of multiple matings.

Here we only measured fecundity and fertility costs; of course, there are other costs that females may experience in mating with heterospecific males. In moving toward calling males, females are subjected to predation and parasitism, and crickets are known to alter their mating behavior in conditions of high predation (Hedrick and Dill 1993). Furthermore acoustically oriented parasitoids might pose risks for females remaining in close proximity to singing males (Cade 1975; Wagner 1996). In addition there is a very severe cost to *G. firmus* females that mate only with heterospecific males – failure to produce progeny. Although failure to mate with any conspecific males is unlikely to occur in the wild, where females are polyandrous and prefer conspecifics (Sakaluk *et al.* 2002; Bretman and Tregenza 2005), in localities where *G. pennsylvanicus* is far more abundant, *G. firmus* females might find few or no conspecific males.

The crickets used in this experiment were all pure species from allopatric populations. In spite of no direct exposure to heterospecifics, these crickets showed strong assortative mating, especially for the cross that produces no hybrid offspring (*G. firmus* female and *G. pennsylvanicus* male). It is possible that behavioral barriers are a byproduct of divergence in allopatry and that the relevant trait differences were already present before secondary contact. It is also possible that secondary character displacement spread from the hybrid zone into the pure species populations adjacent to the hybrid zone.

Studies of mating behavior of crickets within and very far from the hybrid zone are needed to resolve this issue.

A last objective was to determine if *G. pennsylvanicus* males can trigger normal oviposition in *G. firmus* females. Heterospecific males did trigger oviposition in *G. firmus* females, although not to the same extent as conspecific males (Figure 2.4). Females mated only to heterospecific males laid more eggs than virgin females but significantly fewer than females mated to conspecific males. This difference could be due to rapid evolution of accessory gland proteins in *G. firmus* and/or *G. pennsylvanicus* (Andres *et al.* 2006), because transfer from male to female of accessory gland proteins is known to influence oviposition in other insects. Failure to stimulate oviposition may therefore serve as a significant barrier to gene exchange.

ACKNOWLEDGEMENTS

This project was funded by NSF grants DEB-0111613 and DEB-0639904. We thank James Walters for helping with the field collections.

REFERENCES

Alexander, R.D. 1957. The taxonomy of the field crickets of the eastern United States (Orthoptera: Gryllidae: Acheta). Ann. Entomol. Soc. Am. 50: 584-602.

Alexander, R.D. 1968. Life cycle origins, speciation and related phenomena in crickets. Quart. Rev. Biol. 43: 1-41.

Andersson, M. 1994. Sexual Selection. Princeton University Press, Princeton.

Andres, J.A., L.S. Maroja, S.M. Bogdanowicz, W. Swanson, and R.G. Harrison. 2006. Molecular evolution of seminal proteins in field crickets. Mol. Biol. Evol. 23: 1574-1584.

Bailey, R.I., C.D. Thomas, and R.K. Butlin. 2004. Premating barriers to gene exchange and their implications for the structure of a mosaic zone between *Chorthippus brunneus* and *C. jacobsi* (Orthoptera: Acrididae). J. Evol. Biol. 17: 108-119

Bretman, A., and T. Tregenza. 2005. Measuring polyandry in wild populations: a case study using promiscuous crickets. Mol. Ecol. 14: 2169-2179.

Bridle, J.L., C.I. Saldamando, W. Koning, and R.K. Butlin. 2006. Assortative preferences and discrimination by females against hybrid male song in grasshoppers *Chorthippus brunneus* and *Chorthippus jacobsi* (Orthoptera: Acrididae). J. Evol. Biol. 19: 1248-1256.

Broughton, R.E., and R.G. Harrison. 2003. Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. Genetics 163: 1389-1401.

Burpee, D.M., and S.K. Sakaluk. 1993. Repeated matings offset the costs of reproduction in female field crickets. Evol. Ecol. 7: 240-250.

Burnham, K.P., and D.R. Anderson. 1998. Model selection and inference: a practical information-theoretic approach. New York : Springer

Cade, W.H. 1975. Acoustically orienting parasitoids: fly phonotaxis to cricket song. Science 190: 1312-1313.

Crawley, M.J. 1993. GLIM for ecologists. Blackwell, Oxford, UK.

Coyne, J.A. 1992. Genetics and speciation. Nature 355: 511-515.

Coyne, J.A., and H.A. Orr. 2004. *Speciation*. Sunderland, MA: Sinauer Associates, Inc.

Dobzhansky, T. 1937. *Genetics and the Origin of Species*. Columbia Univ. Press, New York.

Dobzhansky, T. 1940. Speciation as a stage in evolutionary divergence. Am. Nat. 74, 312–321.

Doherty, J.A., and M. Storz. 1992. Calling song and selective phonotaxis in field crickets, *Gryllus firmus* and *G. pennsylvanicus* (Orthoptera: Gryllidae). J. Insect Behav. 5: 555-569

Garder, W., E.P. Mulvey, and E.C. Shaw. 1995. Regression analyses of counts and rates: poisson, overdispersed poisson and negative binomial models. Psychol. Bull. 118: 392-404.

Hardy, T.N., and K.C. Shaw. 1983. The role of chemoreception in sex recognition by male crickets *Acheta domesticus* and *Teleogryllus oceanicus*. Physiol. Entomol. 8: 151-166.

Harrison, R.G. 1983. Barriers to gene exchange between closely related cricket species. I. Laboratory hybridization studies. Evolution 37: 245-251.

Harrison, R.G. 1985. Barriers to gene exchange between closely related cricket species. II. Life cycle variation and temporal isolation. Evolution 39: 244-259.

Harrison, R.G. 1986. Pattern and process in a narrow hybrid zone. Heredity 56: 337-349.

Harrison, R.G. 1990. Hybrid zones: windows on evolutionary processes. Pp 69-128 *in* Oxford Surveys in Evolutionary Biology Vol 7 (Futuyma D., and J. Antonovics, Eds) Oxford University Press, New York.

Harrison, R.G., and J. Arnold. 1982. A narrow hybrid zone between closely related cricket species. Evolution 36: 355-552.

Harrison, R.G., and S.M. Bogdanowicz. 1997. Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. Evolution 51: 493-505.

Harrison, R.G., and D.M. Rand. 1989. Mosaic hybrid zones and the nature of species boundaries. Pp 111-133 *in*: Speciation and its consequences (Otte D., and J.A. Endler, Eds). Sinauer, Sunderland, MA.

Hedrick, A.V. 1986. Female preferences for male calling bout duration in a field cricket *Gryllus integer*. Behav. Ecol. Sociobiol 19: 73-77.

Hedrick, A.V., and L.M. Dill. 1993. Mate choice by female crickets is influenced by predation risk. Anim. Behav. 46: 193-196.

Hewitt, G.M. 1988. Hybrid zones - natural laboratories for evolutionary studies. TREE 3: 158-167.

Holzer, B., A. Jacot, and M.W.G. Brinkhof. 2003. Condition dependent signaling affects male sexual attractiveness in field crickets, *Gryllus campestris*. Behav. Ecol. 14: 353-359.

Howard, D.J. 1993. Reinforcement: the origin, dynamics and fate of an evolutionary hypothesis. Pp. 118-142 *in* Hybrid zones and speciation (R.G. Harrison, Ed), Oxford University Press, New York.

Howard, R.W., and Blomquist, G.J. 2005. Ecological, behavioral and biochemical aspects of insect hydrocarbons. Ann. Rev. Entomol. 50: 371-393.

Ivy, T.M., and S.K. Sakaluk. 2005. Polyandry promotes enhanced offspring survival in decorated crickets. Evolution 59: 152-159.

Jiggins, C.D., and J. Mallet. 2000. Bimodal hybrid zones and speciation. TREE 15: 250-255.

Kirkpatrick, M., and V. Ravigne. 2002. Speciation by natural and sexual selection: Models and experiments. Am. Nat. 159: S22-S35.

Lindsey, J.K. 1995. Modelling frequency and count data. Clarendon Press, Oxford, UK.

Liou, L.W., and T. Price. 1994. Speciation by reinforcement of premating isolation. Evolution 48: 1451-1459.

Loher, W., and M. Dambach. 1989. Reproductive behavior. Pp. 43–82 *in* Cricket Behavior and Neurobiology (Huber, F., T.E. Moore, and W. Loher, Eds),. Cornell University Press, Ithaca,

Nagamoto, J., N., Anonuma, and H. Mituhiko. 2005. Discrimination of conspecific individuals via cuticular pheromones by males of the cricket *Gryllus bimaculatus*. Zool. Sci. 22: 1079-1088.

Mayr E. 1963. Animal Species and Evolution. Harvard University Press, Cambridge, MA.

McCullagh, P., and P.A. Nelder. 1989. *Generalized Linear Models*. Chapman and Hall. London, UK.

Mendelson, T.C., and K.L. Shaw. 2002. Genetic and behavioral components of the cryptic species boundary between *Laupala cerasina* and *L. kohalensis* (Orthoptera: Gryllidae). Genetica 116: 301-310.

Mullen, S.P., T.C. Mendelson, C. Schal, and K. Shaw. 2007. Rapid evolution of cuticular hydrocarbons in a species radiation of acoustically diverse Hawaiian crickets (Gryllidae: Trigonidiinae: *Laupala*). Evolution 61: 223-231.

Price, T.D., and M.M. Bouvier. 2002. The evolution of F1 postzygotic incompatibilities in birds. Evolution 56: 2083-2089.

R Development Core Team. 2004. R: A language and environment for statistical computing. Vienna Austria. http://www.R-project.org.

Ramsey, J., H.D. Bradshaw, and D.W. Schemske. 2003. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). Evolution 57: 1520-1534.

Rand, D.M., and R.G. Harrison. 1989. Ecological genetics of a mosaic hybrid zone: mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. Evolution 43: 432-449.

Ross, C.L., and R.G. Harrison. 2002. A fine-scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus*. Evolution 56: 2296-2312.

Ross, C.L., and R.G. Harrison. 2006. Viability selection on overwintering eggs in a field cricket mosaic hybrid zone. Oikos 115: 53-68.

Sakaluk, S.K., J.M. Schaus, A.K. Eggert, W.A. Snedden, and P.L. Brady. 2002. Polyandry and fitness of offspring reared under varying nutritional stress in decorated crickets. Evolution 56: 1999-2007.

Saldamando, C.I., S. Miyaguchi, H. Tatsuta, H. Kishino, J.R. Bridle, and R.K. Butlin. 2005. Inheritance of song and stridulatory peg number divergence between *Chorthippus brunneus* and *C. jacobsi*, two naturally hybridizing grasshopper species (Orthoptera : Acrididae). J. Evol. Biol. 18: 703-712.

Schluter, D. 2001. Ecology and the origin of species. TREE 16: 372-380.

Simmons, L.W. 1986. Female choice in the field cricket, *Gryllus bimaculatus* (De Geer). Anim. Behav. 34:1463–470.

Simmons, L.W. 1988. The calling song of the field cricket *Gryllus bimaculatus* (De Geer): constraints on transmission and its role in intermale competition and female choice. Anim. Behav. 36: 380-394.

Solymar, B.D., and W.H. Cade. 1990. Heritable variation of female mating frequency in field crickets. Behav. Ecol. Sociobiol. 26: 73-76.

Vines, T.H., S.C. Kohler, M. Thiel, T.R. Sands, C.J. MacCallum, N.H. Barton, and B. Nürnberger. 2003. The maintenance of reproductive isolation in a mosaic hybrid zone between the fire bellied toads *Bombina bombina* and *B. variegata*. Evolution 57: 1876-1888.

Wagner, W.E. 1996. Convergent song preferences between female field crickets and acoustically orienting parasitoid flies. Behav. Ecol. 7: 279-285.

Wagner, W.E., R.J. Kelley, K.R. Tucker, and C.J. Harper. 2001. Females receive a life-span benefit from male ejaculates in a field cricket. Evolution 55: 994-1001.

Wells, M.M., and C.S. Henry. 1998. Songs, reproductive isolation, and speciation in cryptic species of insects: A case study using green lacewings. Pp. 217-233 *in* Endless forms: Species and speciation.(Howard, D.J., and S.H. Berlocher, Eds),. Oxford University Press, New York.

Willet, C., M.J. Ford, and R.G. Harrison 1997. Inferences about the origin of a field cricket hybrid zone from a mitochondrial DNA phylogeny. Heredity 79: 484-494.

CHAPTER 3

WOLBACHIA PLAYS NO ROLE IN THE ONE-WAY REPRODUCTIVE INCOMPATIBILITY BETWEEN THE HYBRIDIZING FIELD CRICKETS *GRYLLUS FIRMUS* AND *G. PENNSYLVANICUS*.

ABSTRACT

Wolbachia are cytoplasmically inherited alpha proteobacteria that can cause cytoplasmic incompatibility (CI) in insects. This incompatibility between sperm and egg is evident when uninfected females mate with infected males. *Wolbachia*-driven reproductive incompatibilities are of special interest because they may play a role in speciation. However the presence of *Wolbachia* does not always imply incompatibility. The field crickets *G. firmus* and *G. pennsylvanicus* exhibit a very clear unidirectional incompatibility and have been cited as a possible example of *Wolbachia*-induced CI. Here we conduct curing experiments, intra- and interspecific crosses, cytological examination of *Wolbachia* in testes, and *Wolbachia* quantifications via Real-Time PCR. All of our data strongly suggest that *Wolbachia* are not involved in the reproductive incompatibility between *G. firmus* and *G. pennsylvanicus*.

INTRODUCTION

Wolbachia are cytoplasmically inherited alpha proteobacteria that can infect reproductive tissues of insects and cause reproductive alterations including parthenogenesis (Stouthamer *et al.* 1993), feminization of males (Rousset *et al.* 1992), male killing (Hurst *et al.* 1999) and cytoplasmic incompatibility (Breeuwer *et al.* 1992; O'Neill *et al.* 1992). Surveys have found that 16-76% of insects sampled are infected with *Wolbachia*, with a recent meta-analysis suggesting that 66% of all insect species harbor *Wolbachia* infections at some level (West *et al.* 1998; Jeyaprakash and Hoy 2000; Werren

and Windsor 2000; Hilgenboecker *et al.* 2008). Cytoplasmic incompatibility (CI) is a sperm-egg incompatibility, manifest when uninfected females mate with infected males. Shortly after fertilization, asynchrony in male and female pronuclei development leads to a series of mitotic defects and ultimately to embryonic death (Yen and Barr 1971; Stouthamer *et al.* 1999; Callaini *et al.* 1996; Lassy and Karr 1996; Tram and Sullivan 2002). Although the molecular mechanism of CI is still poorly understood, it appears that *Wolbachia* present inside the testes "modify" the sperm, which must then be "rescued" in the egg by the same *Wolbachia* strain, if successful embryonic development is to occur. *Wolbachia* strains can be classified based on their ability to modify sperm (mod^+ or mod^-) and rescue in eggs ($resc^+$ and $resc^-$) (Werren 1997a).

The effect of CI is unidirectional (usually involving crosses between infected and uninfected individuals) or bidirectional (involving crosses between individuals with different *Wolbachia* types) (Barr 1980; Breeuwer and Werren 1990; O'Neill and Karr 1990). *Wolbachia*-driven reproductive incompatibilities are of special interest because they may play a role in speciation by facilitating the evolution of reproductive isolation between incipient species (Werren 1997b; Telschow *et al.* 2005a; Telschow *et al.* 2005b; Jaenike *et al.* 2006).

The presence of *Wolbachia* does not always imply incompatibility. *Wolbachia* infections exist with no obvious phenotypic effects (Hoffmann *et al.*, 1996). Even in cases where incompatibility is observed, *Wolbachia* may not be the causal agent (Weeks *et al.* 2002). Nuclear genes can also be involved, and in some cases both nuclear-induced and *Wolbachia*-induced incompatibilities are known to play a role (Breeuwer and Werren 1995; Navajas *et al.* 2000; Vala *et al.* 2000). Proving that *Wolbachia* is a causal agent requires curing experiments (treatment with antibiotics) and a rigorous series of crosses between infected and uninfected individuals with the same genetic background.

The hybridizing field crickets *Gryllus firmus* and *G. pennsylvanicus* exhibit a one-way reproductive incompatibility; female *G. firmus* produce no progeny when mated with male *G. pennsylvanicus,* but the reciprocal cross produces viable and fertile offspring (Harrison 1983). The bimodal mosaic hybrid zone formed by these very closely related species is a well-studied model system in speciation research (Harrison 1983, 1985; Harrison and Rand 1989; Willett *et al.* 1997; Ross and Harrison 2002, 2006). Understanding the causes of the one-way incompatibility between *G. firmus* and *G. pennsylvanicus* is important for understanding the evolution of barriers to gene exchange.

Because *Gryllus* species harbor *Wolbachia,* it has been proposed that the bacterial infections are the cause of the one-way reproductive incompatibility (Giordano *et al.* 1997). Giordano *et al.* (1997) argued that *G. pennsylvanicus* was infected whereas *G. firmus* was not, consistent with the pattern expected for *Wolbachia*-induced CI. However, due to incorrect assignment of crickets to species and lack of a perfect correlation between species and infection status, this conclusion was later rejected (Mandel *et al.* 2001). Mandel *et al.* (2001) showed that many *G. firmus* are infected, harboring what they termed the *w*G2 *Wolbachia* strain. Most *G. pennsylvanicus* harbor the *w*G1 strain, but some individuals were doubly infected (*w*G1 and *w*G2), and a few carried only the *w*G2 strain. Extrapolating from these results, Mandel *et al.* (2001) suggested that about 13% of the heterospecific crosses should produce offspring; yet in dozens of crosses observed, not a single one yielded any progeny (Harrison 1983, and R.G. Harrison unpublished data). They concluded that *Wolbachia* is unlike to play a role in the one-way reproductive incompatibility between *G. firmus* and *G. pennsylvanicus*.

Here, we present additional data that argues against a role for *Wolbachia* in reproductive isolation between the cricket species. We conduct curing experiments and intra- and interspecific crosses, use microscopy to examine presence/absence of *Wolbachia* in reproductive tissues, and quantify *Wolbachia* loads in *G. firmus* and *G.*

pennsylvanicus using Real-Time PCR. If *Wolbachia* is responsible for the observed reproductive incompatibility ("the *Wolbachia* hypothesis"), then uninfected (cured) male *G. pennsylvanicus* should be able to sire hybrid progeny when mated with *G. firmus* females. Furthermore, under the "*Wolbachia* hypothesis," *Wolbachia* should be present in testes, and hybrid eggs should be fertilized but later fail to develop. None of these predictions were supported, and we thus reiterate Mandel's *et al.* (2001) conclusion that *Wolbachia* infections are not the cause of the one-way reproductive incompatibility between *G. firmus* and *G. pennsylvanicus*.

MATERIALS AND METHODS

Cricket rearing

We collected late instar *Gryllus firmus* nymphs in Guilford, CT (41°15'; -72°42') and *G. pennsylvanicus* nymphs in Ithaca, NY (42°24'; -76°31'). Both species were collected during August-September 2004. We sorted the crickets by species and maintained five plastic cages (30 × 16 × 9 cm) for each species. Each cage contained five males and five females (total of 50 crickets for each species). Crickets were provided with *ad libitum* food (Purina Cat Chow®), a water vial, cardboard for shelter and a Petri dish of sterilized soil as oviposition substrate. The cages were kept at 25°C, 12:12 light:dark. Oviposition dishes containing eggs were incubated for a maximum of 40 days at 25°C and then placed in a refrigerator at 4°C for 102 days to insure synchronous hatch of nymphs (Harrison 1985).

Hatching started on February 15, 2005, 21 days after eggs were removed from the refrigerator. We divided the offspring from each species into two groups: an antibiotic feeding treatment and an untreated control. Crickets from each group/species were reared in a separate plastic group cage (65 × 45 × 40 cm) with *ad libitum* food (Purina Cat Chow®) and oviposition dishes. The antibiotic treatment group received 0.25%

tetracycline HCl (Sigma) in the water (changed 3 times per week) while the untreated group received pure water. Individuals within each group were allowed to mate freely. Oviposition dishes were treated as described above and hatching began on January 02, 2006. Treatment was continued as in the previous generation. In March 2006 virgin late instar crickets were separated to single-sex plastic cages (30 × 16 × 9 cm), receiving the same treatment as before, until assigned to an experimental cross.

Experimental crosses

Our experimental design for crosses was the most complete possible given the poor performance of our *G. pennsylvanicus* colonies (only 4 untreated and 6 treated males were ultimately available for crossing). Each *G. pennsylvanicus* male was therefore mated to multiple *G. firmus* females. To discover whether *Wolbachia* causes intraspecific CI in *G. firmus*, we also conducted crosses between treated and untreated *G. firmus*.

We abbreviate cross types using 3 letters (e.g., H/tu): The first letter indicates whether the cross is heterospecific or conspecific (H or C), the second letter indicates the male group (treated, t or untreated, u) and the third letter indicates the female group (t or u). Experimental females are always *G. firmus*, thus an H/tt cross involves a treated *G. pennsylvanicus* male and a treated *G. firmus* female while a C/tt cross involves a treated *G. firmus* male and female.

For the heterospecific crosses (H), a treated (t) or untreated (u) *G. pennsylvanicus* male was put with two treated (t) or untreated (u) *G. firmus* females for three days. Females were then removed to individual plastic cages and provided with *ad libitum* food, water and an oviposition dish. The male was placed in an individual cage and reused in subsequent matings. Twenty-six *G. firmus* females were crossed to six treated *G. pennsylvanicus* males: eight females were treated (H/tt) and 18 females were untreated (H/tu). Twenty-six *G. firmus* females were crossed to four untreated males: 10 females

were treated (H/ut) and 16 females were untreated (H/uu). All crosses resulted in spermatophore transfer to the female.

For the conspecific (C) crosses a treated (t) or untreated (u) *G. firmus* male was paired with a treated (t) or untreated (u) *G. firmus* female for three days. The male was then frozen at -80^{o}C for DNA extraction, and the female was placed in an individual plastic cage as described above. We performed 32 crosses using 16 treated and 16 untreated males: five males of each group mated with a treated female (C/tt and C/ut) and 11 males of each group mated with an untreated female (C/tu and C/uu). Again, all females were observed with a spermatophore.

All females used in the crosses were 7-10 days old. Female post-mating lifespan was estimated as days from mating until death. Lifetime fecundity was assessed by counting all eggs laid by each female. Eggs were separated from the oviposition substrate using a series of sieves and counted under a stereoscopic microscope.

Fecundity (number of eggs) and fertility (proportion of eggs hatching) for conspecific and heterospecific crosses were analyzed separately. Female lifespan data were analyzed combining data from conspecific and heterospecific crosses.

Data on conspecific fertility were fitted to a GLM, weighting for fecundity. Conspecific fecundity and female lifespan data were fitted to a GLM with negative binomial errors. Residuals of all the performed GLMs were analyzed by visual inspection and no significant deviations from normality were observed. No outliers were found using Cook's statistics values. The effects of all dependent variables on the response variables were tested using log-likelihood ratio tests comparing the deviance of a model including and excluding the factor being tested. All analyses were performed with R. 2.6.1 (R Development Core Team 2006). All results are shown as (mean ± SD, n).

Real Time PCR

To determine the *Wolbachia* load of each cricket we used TaqMan Real Time PCR. This approach measures the relative number of template molecules for a *Wolbachia*-specific gene compared to template number for a cricket specific gene (Elongation Factor, EF1α). We extracted DNA from whole abdomen of treated and untreated crickets using DNeasy tissue kit (QIAGEN). All DNA extractions were diluted to 10ηg/μl.

Primers and TaqMan fluorescence-labeled probes for real-time PCR assays were designed using Primer Express Software (Applied Biosystems, Foster City, CA). These primers specifically amplify *Gryllus* EF1α (GenBank accession numbers, DQ630925 and DQ630927) and *Wolbachia* ftsZ (GenBank accession numbers, U28195 and U83100, which correspond to the two types of *Wolbachia* found in *G. firmus* and *G. pennsylvanicus*; see Mandel *et al.* 2001). Primers and probes are shown in Table 3.1.

For real-time PCR assays, 2 μL of the extracted genomic DNA template (20ηg) was combined with 900 ηM of each oligonucleotide primer and 250 ηM of the TET-(EF1α) or 6FAM-(ftsZ) and TAMRA- labeled probe in 25 μL of total reaction volume using TaqMan ® Universal PCR Master Mix (Applied Biosystems). TaqMan PCR reactions were mixed in 96-well MicroAmp ™ optical plates (Applied Biosystems). The PCR samples were subjected to 45 cycles of amplification in an ABI 7500 Real Time PCR System (Applied Biosystems) under the following conditions: 50°C for 2 min (uracil N-deglycosylase digest), 95°C for 10 min (AmpliTaq Gold pre-activation), and then 40 cycles of 95°C for 15 sec and 60°C for 1 min. The fluorescence data were analyzed using the Applied Biosystems software. The standard curve was always prepared using a dilution series (up to 10^{-5}) for the same DNA sample; water (in place of DNA) was used as blank. Each individual was assayed at least twice, but because differences in DNA

Table 3.1 Primers and probes for TaqMan real-time PCR for *Gryllus* and *Wolbachia* genes.

Species	Primer or probe	Sequence (5'-3')	position
Gryllus.....	EF1α _F	CTGACCTCCGCAGCAACA	569-587
.	EF1α _R	TTGCCAGTGGTCGAACACA	612-630
	EF1α Probe (TET)	TGGCCAGGCATTCCCTCAGT	591-610
Wolbachia	ftsZ _F	TGAAGAAGTGGATGAAAATGCAAA	729-752
.	ftsZ _R	GCCAGTTGCAAGAACAGAAACTC	800-822
	ftsZ Probe (6FAM)	ACTTTTGATCAGGCGATGGAGGGAAGA	769-795

The nucleotide positions are those reported in GeneBank (accession number DQ630925 for EF1α and U28195 for ftsZ)

concentration between replicates were always <0.01 × (after standardization – see below), we report only the average value for each individual.

Our real-time DNA measurements are relative not absolute; therefore, *Wolbachia* loads reported here are only meaningful for samples in this study. To calculate relative values we assigned an arbitrary DNA quantity for each of the dilutions in the standard curve, from 10^5 in the 1× dilution to 1 in the 10^5× dilution. Based on the Ct (threshold cycle) of each sample and the standard curve, we obtained estimates of EF1α and ftsZ amounts for each individual. We then divided the amount of ftsZ DNA by the amount of EF1α DNA, to generate a normalized value. Finally, a randomly selected untreated individual was chosen to represent a standard 1× *Wolbachia* load, and all other values were adjusted in relation to this standard.

We quantified *Wolbachia* load from a total of 53 crickets. These included parents from all conspecific crosses that failed to produce offspring (C/tt, n=3; C/tu, n=3; C/ut, n=2) as well as from all crosses that produced offspring but were not expected to do so under the hypothesis of *Wolbachia*-induced CI (untreated male and treated female, C/ut, n=3). In addition we quantified all *G. pennsylvanicus* males (n= 5 for treated, n=4 for untreated) and *G. pennsylvanicus* females (n=3 for treated, n=3 for untreated). We also included the following randomly chosen individuals: 5 *G. firmus* treated males, 5 *G. firmus* untreated males, 5 *G. firmus* treated females, and 5 *G. firmus* untreated females. *Wolbachia* loads in one male from an infertile cross (C/tu) and two males from C/ut crosses, as well as one male *G. pennsylvanicus*, could not be measured due to poor quality or unavailable DNA.

We assayed an additional 10 wild caught *G. pennsylvanicus* individuals from Ithaca, NY (5 males and 5 females) to compare *Wolbachia* loads of captive and wild individuals. These crickets were captured as late instar nymphs in August 2007.

To test for differences in *Wolbachia* loads between treatments, sexes, and species, we fitted our data to GLMs with Gamma errors using R. 2.6.1 (R Development Core Team 2004) as explained above (experimental crosses section).

RFLP analysis

Mandel *et al.* (2001) reported 2 different *Wolbachia* strains, *w*G1 (accession number U83100) and *w*G2 (accession number U28195), common to *G. firmus* and *G. pennsylvanicus*. We used an RFLP analysis to determine the strain of *Wolbachia* for each individual assayed with RT-PCR. Universal primers to both strains, ftsZ12F (5'-AAAAATTCAACTTGGTATCAA-3') and ftsZ812R (5'-AGAACAGAAACTCTAACTCTTCC-3'), were used to amplify a short fragment of *Wolbachia* specific ftsZ. These amplifications were carried out in 10µl PCR reactions and contained 3 mM MgCl2, 0.2 mM dNTPs, 50 mM KCl, 20 mM Tris (pH 8.4), 2.5 ηg of each primer, 1 U of Platinum *Taq* DNA polymerase (Invitrogen) and 1µL DNA (20-30ηg). Conditions for thermal cycling were 2 min at 95^{o}C followed by 35 cycles of 50 s at 94°C, 60 s at 55°C and 90 s at 72°C (OmniGene, Hybaid). The resulting 800bp ftsZ fragment is differentially cut by the restriction enzyme *AluI*. We fully digested 3-5µl of PCR product with *AluI* and ran the digests on 2% agarose gels, using a doubly infected individual as a control on each gel. We could unambiguously assign *Wolbachia* strain(s) for all assayed infected individuals.

Microscopy

Gryllus testes and ovaries were dissected from adults in a small petri dish with TBST (50 mM Tris, 150 mM NaCl, 0.1% Tween, 0.05% NaN_3, pH 7.5). Whole testes and ovaries were removed and transferred to a depression slide with TBST. Individual follicles were removed from testes, and ovarioles separated and transferred to 3.7% formaldehyde in TBST for 15-30 minutes followed by three washes in TBST. Tissues were blocked in TBST with 1% BSA for 10 minutes. *Wolbachia* was visualized using an

anti-human hsp60 mouse monoclonal antibody (Sigma), which recognizes *Wolbachia* (Hoerauf *et al.* 2000; McGraw *et al.* 2002). Tissues were incubated in the primary antibody solution (1:500 in TBST, 1% BSA, 2 mg/ml RNaseA and 1:500 of the anti-wsp antibody), for 1 hour at room temperature, followed by three washes with TBST. This was followed by 1 hour at room temperature in 1:500 alexa-flour 488 anti-mouse antibody (Molecular Probes), followed by three washes in TBST. DNA was then stained with either 1 μg/ml DAPI (Molecular Probes) for 5 minutes or with 5 μg/ml Propidium Iodide (Molecular Probes) for 20 minutes, followed by a brief wash in TBST before mounting in ProLong Gold antifade mounting media (Molecular Probes). Images were obtained using a Zeiss Axio-Imager Z1 microscope. We observed both captive and wild-caught untreated crickets as well as 1st and 2nd generation antibiotic treated crickets.

Egg Analysis

Anecdotal evidence indicates that eggs from hybrid crosses (*G. firmus* female × *G. pennsylvanicus*) are smaller and resemble unfertilized eggs. By measuring eggs to the nearest 0.01mm under a dissecting scope (15× magnification), we obtained data on length and width of 10 unfertilized *G. firmus* eggs (virgin females), 10 eggs from crosses within *G. firmus,* and 10 eggs from heterospecific crosses. All eggs were from untreated individuals and were measured during the diapause period (after one month at 4^{o}C).

We also extracted DNA from individual eggs (6 unfertilized, 6 pure *G. firmus* eggs and 6 eggs from crosses between *G. firmus* females and *G. pennsylvanicus* males) using a forensic QIAamp ® DNA micro kit (final dilution in 20μl). All eggs were from untreated females, and DNA was extracted when eggs were in diapause (after one month at 4^{o}C). After estimating the amount of DNA in each sample we used cricket specific microsatellites to test for the presence of maternal and paternal alleles. We also used *Wolbachia* ftsZ primers and cricket specific EF1α primers (as a positive control). PCRs (10μL volume) contained 3 mMMgCl2, 0.2 mMdNTPs, 50 mMKCl, 20 mMTris (pH 8.4),

2.5 ηg of each primer, 1 U of Platinum *Taq* DNA polymerase (Invitrogenl) and 2μL DNA (20-30ηg). PCR amplifications were performed using a thermal cycler (OmniGene, Hybaid) under the following conditions: 40 cycles of 50 s at 94°C,60 s at 52°C and 90 s at 72°C.

RESULTS

Experimental crosses

In heterospecific crosses (*G. pennsylvanicus* male × *G. firmus* female) antibiotic treatment had no effect on whether offspring were produced. All of the heterospecific crosses failed to produce offspring, including those expected to produce offspring under the assumption that CI-inducing *Wolbachia* had been eliminated (Table 3.2). The single hybrid produced (the first ever reported in thousands of hybrid cross observations) came from an H/uu cross, in which both parents were infected as shown by RT-PCR analysis (see below). The fecundity of heterospecific crosses was low (67.9 ± 92.2 eggs, n=52).

On average, conspecific crosses had higher fecundity (373 ± 212 eggs, n=32) and fertility (68.9 ± 67.9 offspring, n=32) than heterospecific crosses, although some females from conspecific crosses (primarily from the treated group) failed to produce any offspring (Table 3.2). The conspecific crosses showed no significant male treatment × female treatment interaction in fecundity, and female treatment alone had a marginally non-significant effect ($F(29, 1)=69$, $P=0.09$) - treated females deposited fewer eggs than untreated females (272 ± 244 eggs, n=10 vs. 419 ± 183 eggs, n=22 respectively). Antibiotic treatment had no significant effect on the percentage of eggs hatching. If *Wolbachia* caused a reproductive incompatibility within species, the C/ut cross should fail to produce offspring (assuming that females were cured); yet this cross was equally fertile.

Antibiotic treatment decreased female post-mating lifespan ($F(80,1)=140.5$, $P<0.0001$), treated females lived for 29.4±14.6 days (n=28) while untreated females lived

Table 3.2 Results of experimental crosses. "Fertility exp?" indicates whether fertility is expected under the *Wolbachia* hypothesis. Results are fecundity (mean eggs and standard deviation), Fertility (mean number of offspring and standard deviation), total number of females in each cross type (n total), number of fecund females (n fecund) and number of fertile females (n fertile). For cross type abbreviations see methods.

	Cross type	Fertility exp?	Mean eggs	SD eggs	Mean offsp	SD offsp	(n) total	(n) fecund	(n) fertile
Conspecific	C/uu	Yes	468.8	159.2	80.2	56.4	(11)	(11)	(11)
	C/tu	Yes	369.4	199.1	69.4	76.5	(11)	(11)	(8)
	C/tt	Yes	293.2	319.1	42.6	58.4	(5)	(5)	(2)
	C/ut	No	250.2	176.4	69.2	92.6	(5)	(5)	(3)
Heterospecific	H/uu	No	86.2	115.2	0.06*	0.25	(16)	(15)	(1)
	H/tu	Yes	64.4	78.1	0	0	(18)	(16)	(0)
	H/tt	Yes	66.5	77.3	0	0	(8)	(8)	(0)
	H/ut	No	46.4	93.5	0	0	(10)	(8)	(0)

* One hybrid offspring was produced in this cross type. Both parents were infected (RT-PCR and RFLP results).

for 43.7±14.5 days (n=56). Unexpectedly mating with conspecifics or heterospecifics also affected female lifespan (F(82, 1)=167.6, $P<0.0001$). Females mated to conspecifics lived longer than females mated to heterospecifics (45.8 ± 14.6 days, n=32 vs. 35.7 ± 15.4 days, n=52 respectively).

Real time PCR

Individual *Wolbachia* loads ranged from 0× to 172×. *Gryllus pennsylvanicus* had a significantly lower *Wolbachia* load than *G. firmus* (F(1, 51) =44.6, $P<0.001$) and females had significantly higher loads than males in both species (F(1, 50) =45.19, $P<0.03$). In addition males responded better to treatment than females (significant sex × treatment interaction (F(1, 46) =5.37, $P<0.03$, Table 3.3). Furthermore, laboratory-reared *G. pennsylvanicus* had significantly lower *Wolbachia* loads than their wild counterparts (F(1, 23) = 45.51, $P<0.001$, Table 3.3).

Only 9 of the 53 assayed individuals showed no evidence of *Wolbachia* infection; 8 of these crickets were treated and one was an untreated *G. pennsylvanicus* female. Of the 9 cured individuals, 2 were *G. pennsylvanicus* (1 male and 1 female) and 7 were *G. firmus* (3 males and 4 females). The cured *G. pennsylvanicus* male was mated to four *G. firmus* females, none of whom produced any offspring. Only one of the cured *G. firmus* females was in a C/ut cross; she mated with a heavily infected male (46× load) and produced 30 offspring (19% hatching success). The other three cured *G. firmus* females were in the C/tt group and these females failed to produce offspring. Two of these cured females were paired with two of the cured males (0× load) and deposited only 1-2 eggs. The third cured female mated with an infected male (0.3× load) and laid 301 eggs, all of which failed to hatch. These three C/tt cured females had very reduced post-mating lifespan: 10 and 11 days for females mated to cured males and 27 days for the female mated to the infected male (compared to 46 days average for females mated to conspecifics). The third cured *G. firmus* male mated with an untreated female (C/tu) and

Table 3.3 *Wolbachia* loads (relative to a standard individual) in *Gryllus firmus* and *G. pennsylvanicus* (mean, standard deviation and number of samples) for treated (t) and untreated (u) males and females.

Species	group	Male			Female		
		mean	SD	(n)	mean	SD	(n)
G. firmus	t	4.2	±7.3	(9)	35.8	±57.6	(13)
	u	22.9	±38.0	(8)	39.6	±26.5	(8)
G. pennsylvanicus	t	0.3	±0.2	(5)	0.8	±0.2	(3)
	u	0.4	±0.2	(4)	0.8	±1.1	(3)
	Wild	5.8	±5.9	(5)	7.5	±9.6	(5)

produced only 12 offspring (3.8% hatching success). The only hybrid offspring (identity confirmed with microsatellites) came from an H/uu cross in which both male and female were infected (0.3× and 16× respectively).

RFLP analysis

We could assign a *Wolbachia* strain to all infected individuals. All experimental *G. pennsylvanicus* had the *w*G1 strain (n=13) and all experimental *G. firmus* had the *w*G2 strain (n=31). However 5 out of 10 wild caught *G. pennsylvanicus* individuals had the *w*G2 type and only 3 had the *w*G1 strain (2 females were uninfected). There were no doubly infected individuals in either experimental or wild groups. An additional sample of 14 *G. pennsylvanicus* from three "pure" populations in New York and Pennsylvania included 9 infected with *w*G1, 4 infected with *w*G2, and one doubly infected.

In the wild caught Ithaca *G. pennsylvanicus* there seems to be a difference in RT-PCR load between *w*G1 and *w*G2 *Wolbachia* strains. The average load for *w*G1 was 1.8× (± 1.2, n=3) and the average for *w*G2 was 12.2× (± 7.0, n=5). Because of the very small sample size we did not conduct any statistical tests.

Microscopy

Wolbachia was easily visualized in ovaries from both *G. firmus* and *G. pennsylvanicus* (Figure 3.1). The overall *Wolbachia* load within ovaries is consistent with the RT-PCR results, with higher *Wolbachia* densities within the ovaries of *G. firmus* than *G. pennsylvanicus* females. Within testes, *Wolbachia* was typically absent from both species (Figure 3.2) for treated, untreated and wild individuals. *Wolbachia* were never seen within the developing spermatocytes, spermatids or surrounding cyst cells. Very rarely, *Wolbachia* could be seen within a single somatic cell (not shown) in the outer follicle epithelium, but this was atypical.

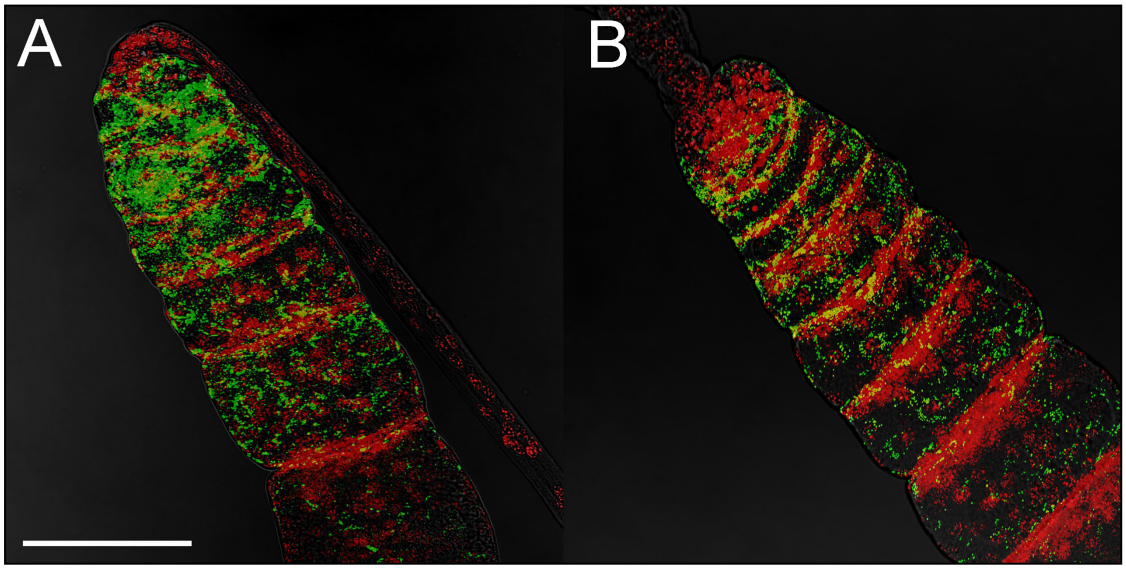


Figure 3.1 *Wolbachia* within developing ovaries of (A) *G. firmus* and (B) *G. pennsylvanicus*.

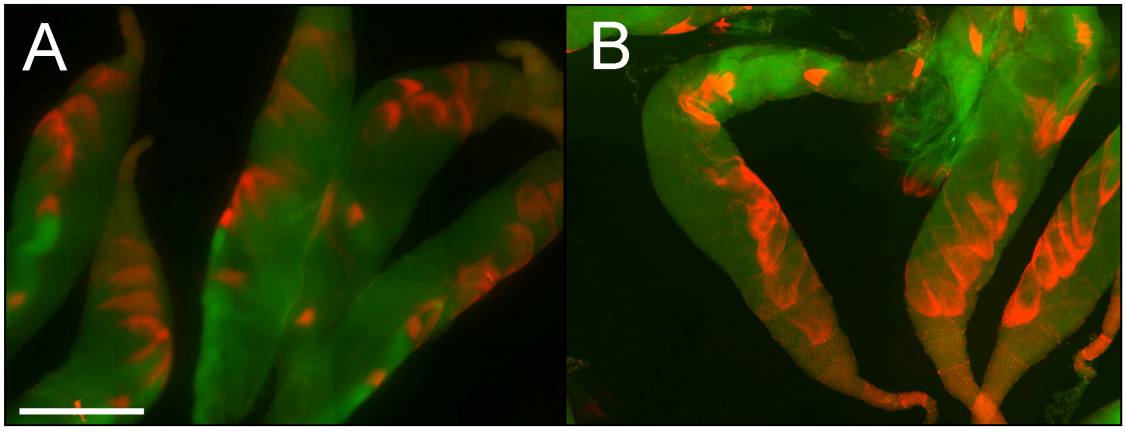


Figure 3.2 *Wolbachia* are absent from testis follicles in both *G. firmus* (A) and *G. pennsylvanicus*. Spermatid nuclei (red) are seen within developing spermatocysts. All FITC staining (green) represents either background staining, or cross-reactivity with spermatid tails.

Egg analyses

We used a one-way ANOVA to test for differences between the mean length and width of unfertilized, hybrid cross, and pure fertilized *G. firmus* eggs (see Figure 3.3). We found no significant difference in mean egg length (unfertilized=3.11mm, hybrid cross=3.08mm, and fertilized *G. firmus*=3.13mm; $F(2, 27)=0.37$, $P>0.5$), but a very significant difference in the mean egg width (unfertilized=0.631mm, hybrid cross=0.641mm, and fertilized *G. firmus*=0.953mm; $F(2, 27)=119.7$, $P<0.001$). The post-hoc Tukey test showed that this difference was between fertilized *G. firmus* eggs and the other two categories. There were no differences between hybrid cross and unfertilized eggs ((U=H) ≠ *Gf*).

All egg DNA extractions had measurable amounts of nucleic acids (mean unfertilized=11ηg/μl, mean hybrid=12ηg/μl and mean fertilized *G firmus*=18ηg/μl); however these measurements probably reflect carrier RNA added during the extractions. We were able to amplify *Gryllus* specific microsatellites and *Wolbachia* ftsZ from all 6 eggs from crosses between *G. firmus* males and females. We also amplified *Wolbachia* ftsZ from 3 of 6 unfertilized eggs and 4 of 6 hybrid cross eggs; however we were unable to amplify cricket specific genes (microsatellites or *EF1-α*) from any of the hybrid cross or unfertilized eggs.

DISCUSSION

Overall we found no evidence that *Wolbachia* infections play a role in the *G. firmus/G. pennsylvanicus* one-way reproductive incompatibility. The original observation (Giordano *et al.* 1997) that *G. pennsylvanicus* is infected and *G. firmus* is not infected is not supported by our data. However, our results do agree with the data of Mandel *et al.* (2001), showing that most individuals in northern populations of both species are infected, that both species harbor strains *w*G1 and *w*G2, but with somewhat different frequencies.

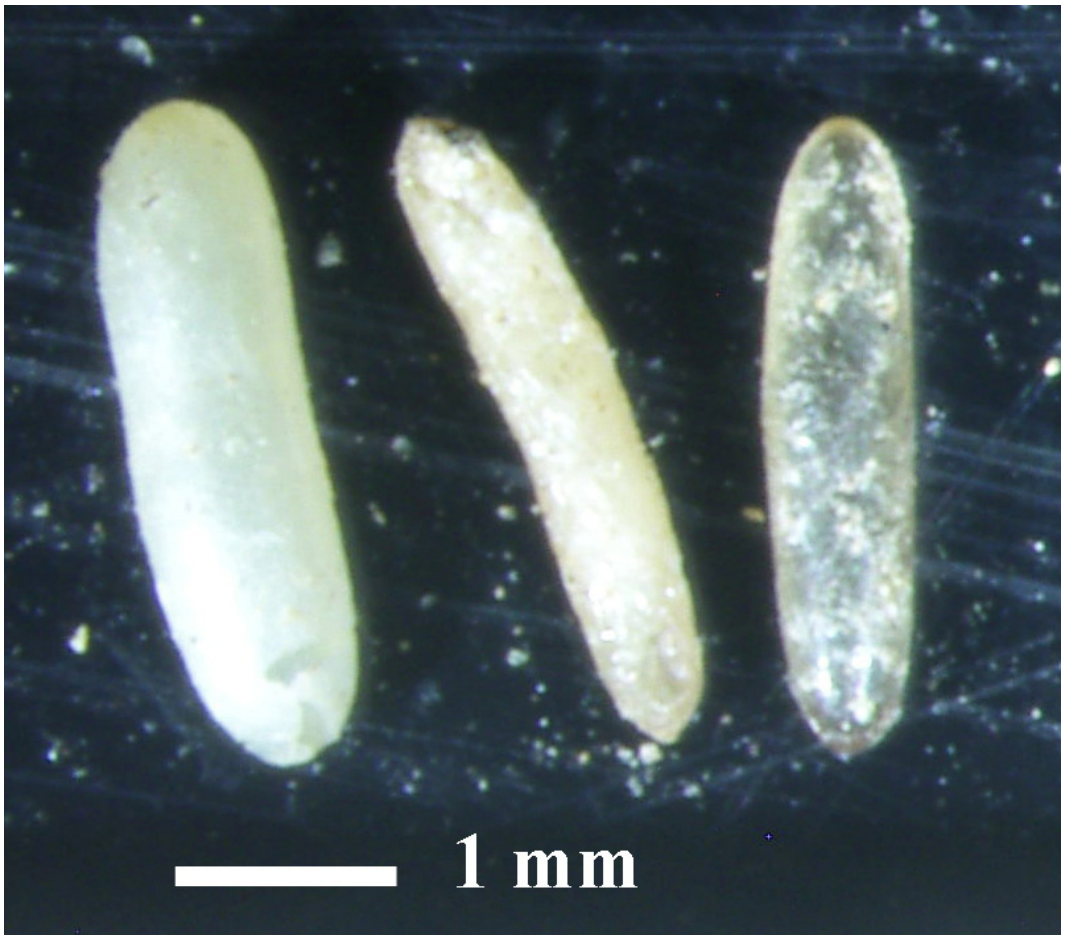


Figure 3.3 Eggs from *G. firmus* females. The first is a typical fertilized *G. firmus* egg from a conspecific cross, the second is an unfertilized egg, and the third is an egg from a hybrid cross. Hybrid cross eggs and unfertilized eggs are always narrower than a fertilized *G. firmus* egg. Color variation is commonly observed in unfertilized and hybrid cross eggs.

Our results show that a completely cured *G. pennsylvanicus* male did not produce offspring when mated with *G. firmus* females, as would be expected if *Wolbachia* was the cause of reproductive incompatibility. Furthermore we did not find *Wolbachia* inside the testes of adult males, although bacteria were present in other tissues. Finally, there is no evidence that eggs from hybrid crosses are fertilized and die later in development as would be expected if *Wolbachia* was responsible for the reproductive incompatibility. *Wolbachia* also does not seem to cause intraspecific CI in crickets; crosses expected to be incompatible (C/ut –untreated males and treated females) produced as many offspring as controls, and a completely cured female was able to produce offspring with a highly infected male. Taken together, these observations provide strong evidence against the hypothesis of *Wolbachia*-induced CI in the field cricket hybrid zone.

Interspecific one way reproductive incompatibility

Bacterial density is associated with the prevalence of *Wolbachia* induced phenotypes, including the expression of CI. Treatments that reduce bacterial densities usually lead to decreases in *Wolbachia* induced phenotypes (Breeuwer and Werren 1993; Hurst *et al.* 2000; Zchori-Fein *et al.* 2000). Thus, if *Wolbachia* was the cause of the observed reproductive incompatibility we would expect a decrease in CI with antibiotic treatment. Although treatment was not 100% effective, males had substantially reduced *Wolbachia* loads, and one *G. pennsylvanicus* male was completely cured (other treated *G. pennsylvanicus* males had loads < 0.5×, Table 3.3). Despite the overall decrease in *Wolbachia* loads (especially in relation to field-collected individuals) and the successful cure, no hybrid crosses produced offspring (except for a single hybrid produced from an infected male and female). Moreover given that higher bacterial densities are associated with increase in incompatibility and possibly higher efficacy in sperm modification and egg rescue (Breeuwer and Werren 1993; Boyle *et al.* 1993; Bressac and Rousset 1993; Poinsot *et al.* 1998; Stouthamer *et al.* 1999), if *Wolbachia* were the cause of reproductive

incompatibility, we would expect higher bacterial loads in *G. pennsylvanicus* (since it should be the species with modified sperm). In contrast to those expectations, *G. firmus* had bacterial loads about 50× those in *G. pennsylvanicus*.

In species with cytoplasmic incompatibility (CI) both bacterial density within testes (Boyle *et al.* 1993; Bressac and Rousset 1993; Giordano *et al.* 1995; Bourtzis *et al.* 1996; Poinsot *et al.* 1998 and Riparbelli *et al.* 2007) and total amount of infected germ cell cysts (Clark *et al.* 2002, 2003) have been implicated in CI expression. Within the testes, CI-inducing (mod^+) *Wolbachia* modify sperm, such that normal embryonic development can occur only if the modification is rescued by an infected ($resc^+$) egg (Werren 1997a). *Wolbachia* strains that neither modify nor rescue sperm (mod^-, $resc^-$) also exist; these strains have no effect on host reproduction (Hoffmann *et al.*, 1996; Veneti *et al.* 2003; Marshall 2004). We did not observe *Wolbachia* inside testes or in the surrounding tissues (Figure 3.2). This absence of *Wolbachia* inside the testes of both *Gryllus* species suggests that these *Wolbachia* strains are unable to modify sperm and cause cytoplasmic incompatibility in these hosts.

Cytoplasmic incompatibility is expected between hybridizing species when one species is infected with CI-causing *Wolbachia* strain and the other species is uninfected or lacks any of the *Wolbachia* types found in the other species. In our experimental crosses, we did not find any doubly infected individuals; all *G. firmus* were infected with the *w*G2 strain and all *G. pennsylvanicus* carried the *w*G1 strain. However wild caught *G. pennsylvanicus* carry either *w*G1 or *w*G2, both strains or are uninfected. Based on our observed ratios of *w*G1-infected, *w*G2-infected, and doubly infected, if *Wolbachia* were the primary cause of the incompatibility between *G. firmus* and *G. pennsylvanicus*, then a large fraction of heterospecific crosses should produce offspring. Contrary to this expectation, in hundreds of heterospecific crosses (Harrison, 1983; Harrison unpublished; Maroja submitted) only one hybrid has ever been produced (reported in this study). This

hybrid came from a cross between untreated crickets, in which the *G. pennsylvanicus* male carried *w*G1 (0.3×load) and *G. firmus* female carried *w*G2 (16×load). The extreme rarity of hybrids from *G. firmus* females suggests the existence of very strong barriers to fertilization. This is corroborated by the observation that mitochondrial DNA introgression across the cricket hybrid zone is always from *G. pennsylvanicus* into *G. firmus* (Harrison *et al.* 1987; Harrison and Bogdanowicz 1997; Ross and Harrison 2002) as would be expected if the only F1 hybrids produced were offspring of *G. pennsylvanicus* females.

Wolbachia induced CI usually leads to early embryonic death or haploid development (Callaini *et al.* 1996; Callaini *et al.* 1997; Duron and Weill 2006; Lassy and Karr 1996; Tram and Sullivan 2002). However, cricket eggs from hybrid crosses resemble unfertilized eggs, both in size (Figure 3.3) and in the failure to provide suitable DNA templates for amplification of *Gryllus* specific microsatellites, while still providing templates for amplification of *Wolbachia* specific genes. If there were early embryos in the eggs from heterospecific crosses then the amount of DNA should be sufficient to allow amplification of *Gryllus* specific microsatellites. It is possible that the DNA has degraded subsequent to the death of the embryo; however in the case we would not expect to be able to amplify *Wolbachia* DNA.

Wolbachia *in conspecific crosses and antibiotic treatment effects*

Wolbachia does not appear to cause intraspecific CI in *G. firmus*. *Wolbachia* is not found in testes of *G. firmus* males (Figure 3.2) and is thus unlikely to modify sperm. Furthermore, our crossing data suggest that the observed infertility (or low fecundity) of some male-female pairs is a result of the antibiotic treatment itself, rather than due to *Wolbachia* presence/absence. If *Wolbachia* caused conspecific CI, and if curing were complete, our C/ut crosses (n=5) should be infertile since an uninfected egg would not be able to rescue modified sperm. The only completely cured female was mated to a highly

infected male (46× load) and had normal fertility. Two C/ut crosses also showed normal fertility (females had reduced *Wolbachia* loads of 5× and 3×). Antibiotic treatment significantly reduces female fecundity independent of the status of the male and probably also explains the infertility of the other two C/ut crosses, as well as other infertile crosses (all of which involved treated individuals). In addition, antibiotic treatment decreased female postmating lifespan. Combined, these observations suggest direct negative effects of antibiotic treatment, effects which confound interpretation of curing experiments in these insects.

Curiously, we also found that mating with conspecific males versus heterospecific males affected female lifespan; females mated to conspecifics lived longer than females mated to heterospecifics. This unexpected result could be related to benefits field cricket females appear to gain from conspecific matings, or an additional unappreciated cost to heterospecific matings (Burpee and Sakaluk 1993; Ivy and Sakaluk 2005; Sakaluk *et al.* 2002; Simmons 1988; Wagner *et al.* 2001). Both costs and benefits might be exaggerated in captive/treated populations. Alternatively, observations of female lifespan differences might be a negative consequence of reduced oviposition rate and accumulation of eggs in females mated to heterospecific males.

Wolbachia *Load*

Total *Wolbachia* loads were consistently much lower in males than females in both species. This may reflect high *Wolbachia* densities in the female germ line and may explain the persistence of *Wolbachia* in the absence of CI. High *Wolbachia* densities in the female germ line likely ensure high rates of *Wolbachia* transmission. Interestingly, wild caught male and female *G. pennsylvanicus* both had much higher *Wolbachia* loads than crickets reared in the lab. The basis of this difference is unclear. Previous reports from *Drosophila* suggest that the phenotypic effects of *Wolbachia* are greater under ideal laboratory conditions compared to either stressed laboratory conditions or wild caught

flies (Ikeda 1970; Hurst *et al.* 2001). Because lab conditions are apparently non-optimal for *G. pennsylvanicus*, like in *Drosophila*, *Wolbachia* loads may be higher in individuals living in more optimal (wild) conditions compared to the sub-optimal (laboratory) conditions.

The data presented here together suggest that *Wolbachia* does not cause CI in *G. pennsylvanicus,* and contrary to previous speculation (Giordano *et al.* 1997). *Wolbachia* does not currently play a role in the incompatibility between *G. pennsylvanicus* and *G. firmus*. Although *Wolbachia* infections have now been described in hundreds of arthropod species, the phenotypic effects of *Wolbachia* infection have been experimentally examined in only a few model organisms. These are restricted to species easily reared in a laboratory environment. A more thorough understanding of effects of *Wolbachia* on hosts will require examination of a wider range of hosts that may not be as amendable to a laboratory environment as traditional model organisms.

Conclusions

Gryllus firmus and *G. pennsylvanicus* are an important model system in the study of speciation and understanding the basis of their reproductive isolation is important for understanding the evolution of barriers to gene exchange. The importance of "infectious speciation" (Coyne 1992) caused by *Wolbachia* bidirectional CI is a subject of current debate (Werren 1997b; Coyne and Orr 2004). There is good evidence that *Wolbachia* infections do play a role in barriers to current gene exchange between species (*Nasonia*, Breeuwer and Werren 1993; Bordenstein *et al.* 2001; *Drosophila*, Shoemaker *et al.* 1999; Jaenike *et al.* 2006). However genic incompatibilities are also involved and it is still unclear whether speciation was caused by *Wolbachia* or by genetic change followed by *Wolbachia* infection. Despite their potential as speciation agents, it seems that the extent to which these endosymbionts play a role in speciation is currently unclear. The field crickets *G. firmus* and *G. pennsylvanicus* exhibit a very clear unidirectional

incompatibility and have been cited as a possible example of *Wolbachia*-induced CI. The data presented here strongly suggest that this is not the case.

ACKNOWLEDGMENTS

We thank Rodman G. Getchell for helping with the RT-PCR, Steve Bogdanowicz for technical advice in all aspects of this study, and Jose A. Andres for statistical advice. This project was supported by NSF grant DEB-0639904 to RGH.

REFERENCES

Barr, A.R. 1980. Cytoplasmic incompatibility in natural populations of a mosquito, *Culex pipiens*. Nature 283: 71-72.

Bordenstein, S.R., F.P. O'Hara, and J.H. Werren. 2001. *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. Science 409: 675-677.

Bourtzis, K., A. Nirgianaki, G. Markakis, and C. Savakis. 1996. *Wolbachia* infection and cytoplasmic incompatibility in *Drosophila* species. Genetics 144: 1063–1073.

Boyle, L., S.L. O'Neill, H.M. Robertson, and T.L. Karr. 1993. Interspecific and intraspecific horizontal transfer of *Wolbachia* in *Drosophila*. Science 260: 1796–1799.

Breeuwer, J.A.J., R. Stouthamer, S.M. Barns, D.A. Pelletier, W.G. Weisburg, and J.H. Werren. 1992. Phylogeny of cytoplasmic incompatibility microorganisms in the parasitoid wasp genus *Nasonia* based on 16S ribosomal DNA sequences. Insect Mol Biol 1: 25-36.

Breeuwer, J.A.J., and J.H. Werren. 1990. Microorganisms associated with chromosome destruction and reproductive isolation between two insect species. Nature 346: 558-560.

Breeuwer, J.A.J., and J.H. Werren. 1993. Microorganisms associated with chromosome destruction and reproductive isolation between two insect species. Nature 346: 558-560.

Breeuwer, J.A.J., and J.H. Werren. 1995. Hybrid breakdown between two haploid species: the role of nuclear and cytoplasmic genes. Evolution 49: 705-717.

Bressac, C., and F. Rousset. 1993. The reproductive incompatibility system in *Drosophila simulans*: DAPI-staining analysis of the *Wolbachia* symbionts in sperm cysts. J Invert Pathol 61: 226–230.

Burpee, D.M., and S.K. Sakaluk 1993. Repeated matings offset the costs of reproduction in female field crickets. Evol Ecol 7:240-250.

Callaini, G., R. Dallai, and M.G. Riparbelli. 1997. *Wolbachia*-induced delay of paternal chromatin condensation does not prevent maternal chromosomes from entering anaphase in incompatible crosses of *Drosophila simulans*. J. Cell Sci 110: 271–280.

Callaini, G.M., G. Riparbelli, R. Giordano, and R. Dallai. 1996. Mitotic defects associated with cytoplasmic incompatibility in *Drosophila simulans*. J. Invert Pathol 67: 55–64

Clark, M.E., Z. Veneti, K. Bourtzis, and T.L. Karr. 2002. The distribution and proliferation of the intracellular bacteria *Wolbachia* during spermatogenesis in *Drosophila*. Mech Dev 111: 3–15.

Clark, M.E., Z. Veneti, K. Bourtzis, and T.L. Karr. 2003. *Wolbachia* distribution and cytoplasmic incompatibility during sperm development: the cyst as the basic cellular unit of CI expression. Mech Dev 120: 185–198.

Coyne, J.A. 1992. Genetics and speciation. Nature 355: 511-515.

Coyne, JA, and H.A. Orr. 2004. *Speciation*. Sunderland, MA: Sinauer Associates.

Duron, O., and M. Weill. 2006. *Wolbachia* infection influences the development of *Culex pipiens* embryos in incompatible crosses. Heredity 96: 493-500.

Giordano, R., J.J. Jackson, and H.M. Robertson. 1997. The role of *Wolbachia* bacteria in reproductive incompatibilities and hybrid zones of *Diabrotica* beetles and *Gryllus* crickets. Proc Natl Acad Sci USA 94:11439-11444.

Giordano, R. S.L. O'Neil, and H.M. Robertson. 1995. *Wolbachia* infections and the expression of cytoplasmic incompatibility in *Drosophila sechellia* and *D. mauritiana*. Genetics 140: 1307-1317.

Harrison, R.G. 1983. Barriers to gene exchange between closely related cricket species. I. Laboratory hybridization studies. Evolution 37: 245-251.

Harrison, R.G. 1985. Barriers to gene exchange between closely related cricket species. II. Life cycle variation and temporal isolation. Evolution 39: 244-259.

Harrison, R.G., and S.M. Bogdanowicz. 1997. Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. Evolution 51: 493-505.

Harrison, R.G., and D.M. Rand. 1989. Mosaic hybrid zones and the nature of species boundaries. Pp 111-133 *in*: Speciation and its consequences (Otte, D., and J.A. Endler, Eds). Sinauer, Sunderland, MA.

Harrison, R.G., D.M. Rand, and W.C. Wheeler. 1987. Mitochondrial DNA variation in field crickets across a narrow hybrid zone. Mol Biol Evol 4: 144-158.

Hilgenboecker K., P. Hammerstein, P. Schlattmann, A. Telschow, and J.H. Werren. 2008. Incidence and Prevalence of *Wolbachia*. Molecular Ecology (In Press).

Hoerauf, A., L. Volkmann, C. Hamelmann, O. Adjei, I.B. Autenrieth, B. Fleischer, and D.W. Buttner. 2000. Endosymbiotic bacteria in worms as targets for a novel chemotherapy in filariasis. Lancet 355:1242-1243.

Hoffmann, A.A., D. Clancy, and J. Duncan. 1996. Naturally-occurring *Wolbachia* infection in *Drosophila simulans* that does not cause cytoplasmic incompatibility. Heredity 76: 1-8.

Hurst, G.D., F.M. Jiggins, and S.J.W. Robinson. 2001. What causes inefficient transmission of male-killing *Wolbachia* in *Drosophila*? Heredity, 87: 220-226.

Hurst, G.D., A.P. Johnson, J.H. Schulenburg, and Y. Fuyama. 2000. Male-killing *Wolbachia* in *Drosophila*: a temperature-sensitive trait with a threshold bacterial density. Genetics 156: 699-709.

Hurst, G.D., F.M. Jiggins, J.H.G. Schulenburg, D. Bertrand, S.A. West, I.I. Goriacheva, . A. Zakharov, J.H. Werren, R. Stouthamer, and M.E.N. Majerus.

1999. Male-killing Wolbachia in two species of insect. Proc R Soc Lond B 266: 735-740.

Ikeda, H. 1970. The cytoplasmically-inherited 'sex-ratio' condition in natural and experimental populations of *Drosophila bifasciata*. Genetics, 65: 311-333.

Ivy, T.M., and S.K. Sakaluk. 2005. Polyandry promotes enhanced offspring survival in decorated crickets. Evolution 59: 152-159.

Jaenike, J., K.A. Dyer, C. Cornish, and M.S. Minhas. 2006. Asymmetrical reinforcement and *Wolbachia* infection in *Drosophila*. PLoS Biol. 4:1852-1862.

Jeyaprakash, A, and M.A. Hoy. 2000. Long PCR improves *Wolbachia* DNA amplification: wsp sequences found in 76% of sixty-three arthropod species. Insect Mol Biol 9: 393-405.

Lassy, C.W., and T.L. Karr. 1996. Cytological analysis of fertilization and early embryonic development in incompatible crosses of *Drosophila simulans*. Mech. Dev. 57: 47–58.

Mandel, M.J., C. Ross, and R.G. Harrison. 2001. Do *Wolbachia* play a role in unidirectional incompatibilities in a field cricket hybrid zone? Mol Ecol 10: 703-709.

Marshall, J.L. 2004. The *Allonemobius*-*Wolbachia* host-endosymbiont system: evidence for rapid speciation and against reproductive isolation driven by cytoplasmic incompatibility. Evolution 58: 2409-2425.

McGraw, E.A., D.J. Merritt, N.J. Droller, and S.L. O'Neill. 2002. *Wolbachia* density and virulence attenuation after transfer into a novel host. Proc Natl Acad Sci USA 99: 2918-2923.

Navajas, M., A. Tsagkarakov, J. Lagnel, and M.-J. Perrot-Minnot. 2000. Genetic differentiation in *Tetranychus urticae* (Acari: Tetranychidae): polymorphism, host races or sibling species? Exp. Appl. Acarol. 24: 365-376.

O'Neil, S.L., R. Giordano, A.M.E. Colbert, T.L. Karr, and H.M. Robertson. 1992. 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. Proc Natl Acad Sci USA 89: 2699-2702.

O'Neill, S.O., and T.L. Karr. 1990. Bidirectional cytoplasmic incompatibility between conspecific populations of *Drosophila simulans*. Nature 348: 178–180.

Poinsot, D., K. Bourtzis, G. Markakis, C. Savakis, and H. Merlot. 1998. *Wolbachia* transfer from *Drosophila melanogaster* into *D. simulans*: host effect and cytoplasmic incompatibility relationships. Genetics 150: 227–237.

Riparbelli, M.G., R. Giordano, and G. Callaini. 2007. Effects of *Wolbachia* on sperm maturation and architecture in *Drosophila simulans* Riverside. Mech Dev 124: 699-714.

Ross, C.L., and R.G. Harrison. 2002. A fine-scale spatial analysis og the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus*. Evolution 56: 2296-2312.

Ross, C.L., and R.G. Harrison. 2006. Viability selection on overwintering eggs in a field cricket mosaic hybrid zone. Oikos 115: 53-68.

Rousset, F., D. Bouchon, B. Pintureau, P. Juchault, and M. Solignac. 1992. *Wolbachia* endosymbionts responsible for various alterations of sexuality in arthropods. Proc R Soc London B 250: 91-98.

Sakaluk, S.K, J.M. Schaus, A.K. Eggert, W.A. Snedden, and P.L. Brady. 2002. Polyandry and fitness of offspring reared under varying nutritional stress in decorated crickets. Evolution 56: 1999-2007.

Simmons, L.W. 1988. The calling song of the field cricket *Gryllus bimaculatus* (De Geer): constrains on transmission and its role in intermale competition and female choice. Anim Behav 36: 380-394.

Stouthamer, R., J.A.J. Breeuwer, and D.D. Hurst. 1999. *Wolbachia pipientis*: Microbial manipulator of arthropod reproduction. Ann Rev Microbiol 53: 71-102.

Stouthamer, R., J.A.J. Breeuwer, R.F. Luck, and J.H. Werren. 1993. Molecular identification of microorganisms associated with parthenogenesis. Nature 361: 66-68.

Telschow, A., P. Hammerstein, and J.H. Werren. 2005a. The effect of *Wolbachia* versus genetic incompatibilities on reinforcement and speciation. Evolution 59:1607-1619.

Telschow, A., Yamamura, N., and J.H. Werren. 2005b. Bidirectional cytoplasmic incompatibility and the stable coexistence of two *Wolbachia* strains in parapatric host populations. J Theor Biol 235:265-274.

Tram, U., and W. Sullivan. 2002. Role of delayed nuclear envelope breakdown and mitosis in *Wolbachia*-induced cytoplasmic incompatibility. Science 296: 1124–1126

Vala, F., J.A.J. Breeuwer, and M.W. Sabelis. 2000. *Wolbachia* induced 'hybrid breakdown' in the two-spotted spider mite *Tetranychus urticae* Koch. Proc R Soc London B 267: 1931-1937.

Veneti, Z., M.E. Clark, S. Zabalou, T.L. Karr, C. Savakis, and K. Bourtzis. 2003. Cytoplasmic Incompatibility and Sperm Cyst Infection in Different *Drosophila-Wolbachia* Associations. Genetics 164: 545-552.

Wagner, W.E., R.J. Kelley, K.R. Tucker, and C.J. Harper. 2001. Females receive a life-span benefit from male ejaculates in a field cricket. Evolution 55: 994-1001.

Weeks, A., K.T. Reynolds, and A.A. Hoffmann. 2002. *Wolbachia* dynamics and host effects: what has (and has not) been demonstrated? TREE 17: 257-262.

Werren, J.H. 1997a. Biology of Wolbachia. Annu Rev Entomol 42: 587-609.

Werren, J.H. 1997b. *Wolbachia* and speciation. Pp 245-260 *in* Endless Species and Speciation (Howard, D. and S. Berlocher, Eds). Oxford University Press, New York.

Werren, J.H., and D.M. Windsor. 2000. *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? Proc R Soc Lond B Biol Sci 267: 1277-1285.

West, S.A., J.M. Cook, J.H. Werren, and H.C. Godfray. 1998. *Wolbachia* in two insect host-parasitoid communities. Mol Ecol 7: 1457-1465.

Willett, C., M.J. Ford, and R.G. Harrison. 1997. Inferences about the origin of a field cricket hybrid zone from a mitochondrial DNA phylogeny. Heredity 79: 484-494.

Yen, J.H., and A.R. Barr. 1971. New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens*. Nature 232: 657-658.

Zchori-Fein, E., Y. Gottlieb, and M. Coll. 2000. *Wolbachia* density and host fitness components in *Muscidifurax uniraptor* (Hymenoptera: pteromalidae). J Invertebr Pathol 75: 267-272.

CHAPTER 4

GENEALOGICAL DISCORDANCE AND PATTERNS OF INTROGRESSION AND SELECTION ACROSS A CRICKET HYBRID ZONE.

ABSTRACT

In recently diverged species, ancestral polymorphism and introgression can cause incongruence between gene trees and species trees. In the face of hybridization only genomic regions that cannot cross the species boundaries will exhibit reciprocal monophyly. These regions, usually evolving rapidly under selection, may be few in number, but are important for the maintenance of species boundaries. In animals with internal fertilization, genes encoding seminal protein are candidate barrier genes. Recently diverged hybridizing species such as the field crickets *Gryllus firmus* and *G. pennsylvanicus,* offer excellent opportunity to investigate the origins of barriers to gene exchange. These recently diverged cricket species form a well-characterized hybrid zone in the eastern United States, and share ancestral polymorphisms across the genome. We analyzed DNA sequence divergence for seminal protein loci, housekeeping loci, and mtDNA, using a combination of analytical approaches and extensive sampling across both species and the hybrid zone. We report discordant genealogical patterns and differential introgression rates across the genome. The most dramatic outliers, showing near zero introgression and more structured species trees, are also the only two seminal protein loci under selection. These are candidate barrier genes with possible reproductive functions in field crickets. We also use the genealogical data to examine the demographic history of the field crickets and the current structure of the hybrid zone.

INTRODUCTION

Introgressive hybridization is now recognized as an important process in evolution (Arnold 1997) and has been documented in a variety of animal species (Wang *et al.* 1997; Besansky *et al.* 2003; Machado *et al.* 2002; Grant *et al.* 2004; Seehausen 2004; Putnam *et al.* 2007; Kronforst 2008), as well as in plants and prokaryotes (Grant 1981; Rieseberg 1997; Jain *et al.* 2002). However alleles at some loci are not "free" to move across species boundaries. In some genomic regions introgression will be limited or prevented by incompatibilities, resulting in a semi-permeable species boundary (Barton and Hewitt 1981; Harrison 1990; Wu 2001). These genomic regions might be very few, but they are essential for the maintenance of species boundaries in the face of hybridization.

Gene introgression violates assumptions of the basic bifurcating model of species divergence and, together with shared ancestral polymorphism, can cause incongruence between gene trees and species trees (Neigel and Avise 1986; Hudson 1992; Nichols 2001). Thus the genome of recently diverged/hybridizing species will be a mosaic of different genealogical histories (Ting *et al.* 2000; Hudson and Coyne 2002; Machado and Hey 2003; Dopman *et al.* 2005; Broughton and Harrison 2003; Andrés *et al.* in press). In these species only genomic regions that cannot cross species boundaries or that have experienced recent selective sweeps will exhibit reciprocal monophyly (exclusivity). Thus, searching the genome for regions showing lack of introgression and/or species monophyly can potentially reveal so called “speciation” or “barrier” genes (Rieseberg *et al.* 1999; Wu 2001; Dopman *et al.* 2005; Payseur and Nachman 2005; Noor and Feder 2006).

Barrier genes are often involved in reproductive incompatibilities and may be evolving rapidly under selection. Recently, evolutionary geneticists have made great progress in the identification of barrier/speciation genes in model organisms. For example several major effect genes, most of them under selection, have been described: *Xmrk-2*

causes inviability in hybrid platyfishes (Wittbrodt *et al.* 1999), *OdsH* and *JYAlpha* cause hybrid male sterility in *Drosophila (*Ting *et al.* 1998; Wu and Ting 2004; Masly *et al.* 2006) and *Hmr, Nup96* and *Lhr* cause hybrid inviability in *Drosophila* (Barbash *et al.* 2003; Presgraves *et al.* 2003; Brideau *et al.* 2006). Analyzing patterns of introgression across the *Mus musculus* and *M. domesticus* hybrid zone, Payseur and Nachman (2005) identified seven candidate barrier genes that showed high rates of protein evolution and male limited expression. In organisms with more limited genetic resources, finding barrier genes has been more challenging. Nonetheless, candidate barrier genes have been identified using statistical analysis of hybrid zones (e.g. Riesenberg *et al.* 1999; Grahame *et al.* 2006), analysis of gene genealogies (e.g. Dopman *et al.* 2005; Andrés *et al.* in press), population genetics (e.g. Vasemagi *et al.* 2005; Nosil *et al.* 2008) and coalescent based approaches (e.g. Putnam *et al.* 2007).

In animals with internal fertilization, genes encoding seminal proteins represent a class of rapidly evolving and often positively selected genes that are potential candidate barrier genes (Swanson and Vacquier 2002). Seminal proteins are transferred to females along with sperm during copulation and play an important role in reproductive interactions and potentially in the evolution of reproductive isolation. For example, in *Drosophila*, seminal proteins have been shown to influence female physiology and behavior including oogenesis, ovulation, oviposition, sperm storage and remating rates (e.g. Harshman and Prout 1994; Herndon and Wolfner 1995; Wolfner 1997; Neubaum and Wolfner 1999; Tram and Wolfner 1999). In both insects and primates, some of these proteins exhibit a clear signature of positive selection (Clark *et al.* 2006) and may be an important component of reproductive isolation during the early stages of the speciation process (Andrés and Arnqvist 2001).

Recently diverged species that continue to hybridize offer excellent opportunity to investigate the origins of barriers to gene exchange. The field crickets *Gryllus firmus* and

G. pennsylvanicus are very closely related (<0.5% mtDNA divergence – Willett *et al.* 1997), come into contact in a well-characterized hybrid zone in eastern North America (Harrison and Bogdanowicz 1997; Ross and Harrison 2002), do not form exclusive groups and share ancestral polymorphisms at many loci across the genome (Harrison 1979; Broughton and Harrison 2003). Although morphologically similar (Alexander 1957), these crickets have clearly diverged in ecology (Rand and Harrison 1989; Ross and Harrison 2002) and behavior (Harrison and Rand 1989; Doherty and Storz 1992; Maroja *et al.* chapter 2). Furthermore they exhibit a clear reproductive incompatibility: *G. pennsylvanicus* males do not trigger normal oviposition in *G. firmus* females (Harrison 1983; Maroja *et al.* chapter 2), and the eggs produced fail to develop and are indistinguishable from unfertilized eggs (Harrison 1983; Maroja *et al.* chapter 2 and chapter 3). If seminal proteins are involved in these reproductive barriers, they may show the signature of restricted introgression and species exclusivity (monophyly).

In an effort to identify proteins that might be responsible for reproductive isolation, Andrés *et al.* (2006) characterized accessory gland genes from *G. firmus* and *G. pennsylvanicus*, many of which are rapidly evolving and under selection (see also Braswell *et al.* 2006). Subsequent proteomic analyses provided unambiguous identification of seminal proteins (Andrés *et al.* in press). Here we generate genealogies for six seminal protein loci, three "housekeeping" loci and mtDNA, using extensive population sampling across both field cricket species and the hybrid zone. Using a combination of analytical approaches, we show that introgression varies strikingly across the genome. Furthermore, two nuclear loci that show a pattern consistent with absence of introgression encode seminal proteins that are under positive selection. We also use the genealogical data to examine the demographic history of the field crickets and the current structure of the hybrid zone.

MATERIALS AND METHODS

Population sampling

We collected *Gryllus firmus* and *G. pennsylvanicus* from 14 populations (Figure 4.1, Table 4.1). Six of these populations represent "pure" species: Guilford, CT (GUI, n= 6), Tom's River, NJ (TRI, n= 5) and Parksley, VA (PAR, n= 6) represent "pure" *G. firmus* populations while Ithaca, NY (ITH, n= 6), Scranton, PA (SCR, n= 4) and State College, PA (SCO, n= 5) represent "pure" *G. pennsylvanicus* populations. We confirmed the non-hybrid population status with phenotypic measurements. In addition to *G. firmus* and *G. pennsylvanicus*, we also sampled *G. rubens* from Durham, NC (n=5) and from Roanoke, VA (n=2) and *G. bimaculatus* from a colony maintained by the Hoy lab at Cornell University (n=2).

Gene sequencing and allele inference

In this paper we focus on one mitochondrial DNA gene (*mtDNA*), Cytochrome Oxidase I (including part of the adjacent t-RNA), and nine nuclear autosomal genes, all of which were isolated from a field cricket male accessory gland cDNA library (Andrés *et al.* 2006). The nuclear genes include *Hexokinase* (*Hex*), *Elongation Factor* 1-α (*EF*1-α), *Guanylate Kinase*-1 (*GuKc*) and six anonymous proteins (*AG-0005F*, *AG-0032F*, *AG-0090F*, *AG-0211F*, *AG-0254P* and *AG-0334P*). *Hex* is an enzyme that phosphorylates hexose, participating in the first step of the glycolytic pathway. *EF1-α* is an essential component of the eukaryotic transcriptional apparatus catalyzing the transfer of aminoacyl-transfer RNA to the ribosome. *GuKc* catalyses the ATP-dependent phosphorylation of GMP into GDP. These genes do not show male biased expression, are not secreted (based on absence of a signal peptide), and likely do not function in cricket reproduction. The genes encoding these proteins are therefore considered "housekeeping genes". Of the six unidentified proteins, *AG-0254P* is secreted (has a signal peptide) and possibly has an unknown binding function, because it exhibits similarity to OS-D

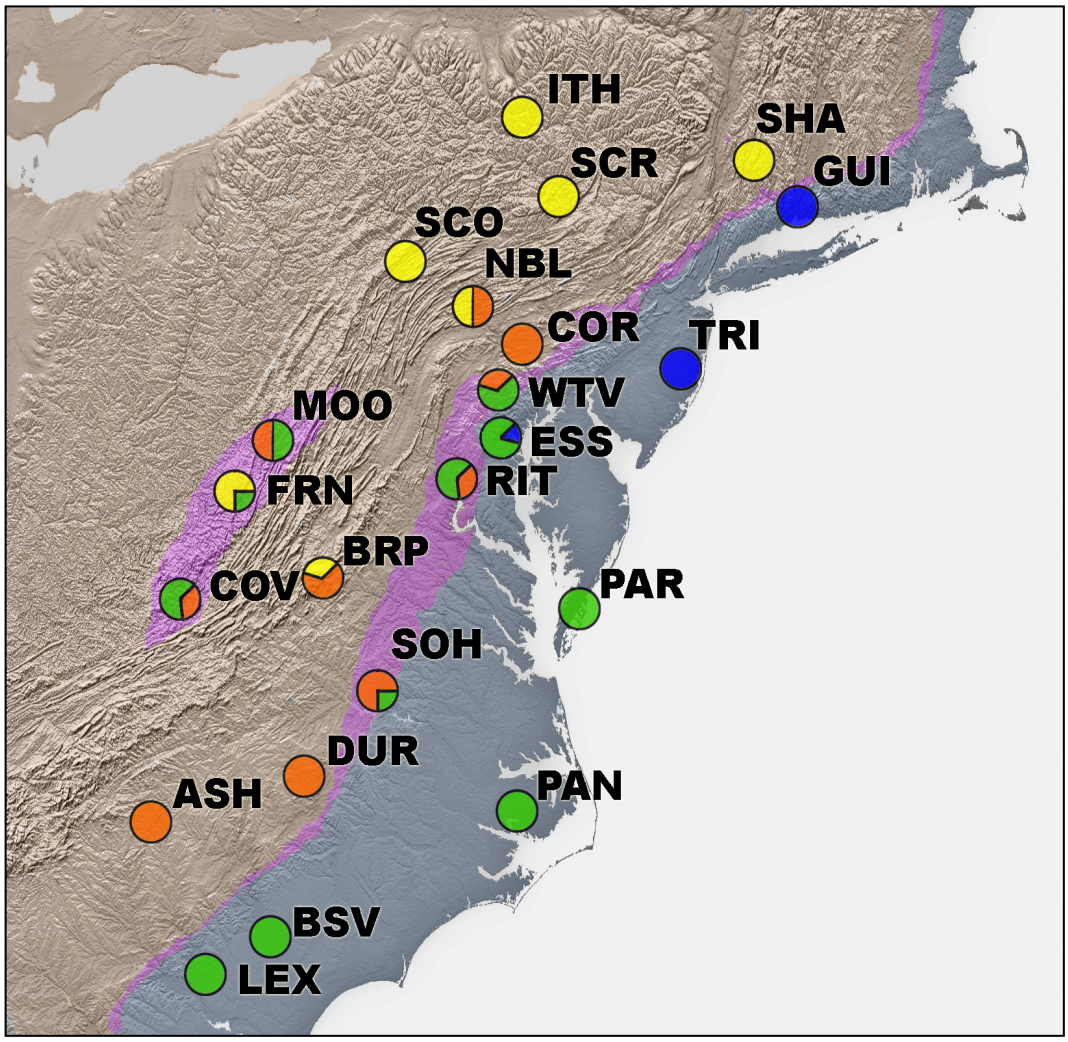


Figure 4.1 Collection localities. Seven populations from Willett *et al.* (1997) were included the mtDNA analysis (ASH, BRP, BSV, COR, LEX, PAN, SHA, WTV). Population colors represent clade affiliation based on the mtDNA phylogeny. Yellow and orange represent the Northern and Southern *G. pennsylvanicus* clades; blue and green represent the Northern and Southern *G. firmus* clades. Blue shaded background shows G. firmus distribution, brown shaded background shows *G. pennsylvanicus* distribution and purple areas show hybrid populations.

Table 4.1 Sampled populations; pure *G. pennsylvanicus* in bold and pure *G. firmus* in italics.

Population	Abr.	Latitude (N)	Longitude (W)	Elevation	n[a]	total bp[b]	π ± SD[c]	
Ithaca, NY	ITH	42 26 01	76 29 59	250m	10.4	7162	0.0063	0.0039
Scranton, PA	SCR	41 24 25	75 35 46	397m	7.4	7301	0.0060	0.0039
Guilford, CT	GUI	41 16 48	72 42 02	0m	10	6956	0.0074	0.0059
State College, PA. . .	SCO	40 47 59	77 52 05	371m	9.3	6675	0.0064	0.0047
New Bloomfield, PA	NBL	40 28 24	77 07 50	379m	10.2	6938	0.0073	0.0047
Tom's River, NJ	TRI	39 45 00	74 11 33	0m	7.8	7133	0.0064	0.0036
Essex, MD	ESS	39 18 20	76 28 46	0m	8.5	7074	0.0084	0.0046
Moorefield, WV	MOO	39 04 09	78 55 58	285m	8.1	7058	0.0091	0.0046
Ritchie, MD	RIT	38 52 07	76 51 01	0m	9.8	7095	0.0073	0.0041
Franklin, WV	FRN	38 39 20	79 19 59	551m	7.2	7548	0.0080	0.0045
Covington, VA	COV	38 00 50	78 28 21	354m	9.6	7083	0.0088	0.0056
Parksley, VA	PAR	37 45 58	75 36 00	0m	10.8	7118	0.0078	0.0054
South Hill, VA	SOH	36 45 07	78 06 09	116m	7.4	6713	0.0079	0.0068
Durham, NC	DUR	36 03 23	79 04 45	159m	0.67*	2717	0.0025	0.0036

[a] Average number of haplotypes sequenced.

[b] Total number of base pair sequenced across all loci.

[c] Average π and SD across all 10 loci.

* Only 4 haplotypes available for *AG-0005* and *mtDNA*

chemosensory protein. The other five seminal proteins are biochemically uncharacterized, are secreted and/or show male biased expression (Andrés *et al.* 2006, Table 4.2). Through proteomic analysis, the proteins encoded by *AG-0005F*, *AG-0090F* and *AG-0334P* are known to be present in the spermatophore (Andrés *et al.* in press, Table 4.2). Locus specific information on primer sequence, number of sequenced base pairs, total number of coding nucleotides, total number of variable sites, male expression bias and whether protein is secreted can be found on Table 4.2. All sequences have been deposited in GenBank (XXX-XXX).

Genomic DNA was isolated from leg muscle tissue using the DNeasy tissue kit (QIAGEN). Locus specific primers (Table 4.2) were used to PCR amplify each of the 10 loci. PCR reactions (10 μl volume) contained 3mM $MgCl_2$, 0.2 mM dNTPs, 50mM KCl, 20mM Tris (pH 8.4), 2.5ng of each primer and 1 U of *Taq* DNA polymerase (Gibco-BRL) and 1 μL DNA. PCR amplifications were performed using a thermal cycler (OmniGene, Hybaid) under the following touchdown conditions: 10 cycles of 50 s at 95°C, 60 s at 65-55°C (decreasing 1°C per cycle) and 90 s at 72°C followed by 30 cycles of 50 s at 95°C, 60 s at 55°C and 90 s at 72°C. All genes were sequenced in both directions. Sequences were aligned in SeqMan (DNAstar) and SNPs were identified by visual inspection; only high quality traces were considered. Individual haplotypes were reconstructed using the PHASE algorithm (Stevens *et al.* 2001). Excluding autapomorphies, all haplotype identifications had posterior probabilities greater than 0.8.

We sequenced at least 4 individuals (8 haplotypes) from each population for each locus (average number of haplotypes per locus/per population is shown in Table 4.1). Durham, NC population (DUR) had only two crickets available and was sequenced only for *mtDNA* and *AG-0005F*.

Table 4.2 Locus specific information, including primers used, whether gene products are secreted, have male biased gene expression.

Loci	5'-3'Primers forward	5'-3'Primers reverse	Secr[a]	♂ bias[b]	total[c]	cod[d]	n[e]	var[f]
mtDNA. . . .	ACCCCCATCATTAACCCTTTTA	GAGACCATTACTTGCTTTCAGTCATCT*	No	No	1889	1182	70**	187
EF1-α.	CGAAATCGCCTAACAAACATAACA	AATCCTTTCCTCTTGCGTGTG	No	No	727	372	132	61
GuKc.	GCTGCTAATCGCGGAAGTGC	GCTGCCTTGCTTGTGCCATAC	No	No	434	156	126	17
Hex.	AATGGGGAGCTTTCGGAGAT	CATTGGCACAGTTTTGGTCAG	No	No	460	282	126	18
AG-0005F. .	GATGAGGCTGCTGGTCGTCGTG	GTGGTTAGCAGGGGCGTGATGGTT	Sp	Yes	911	911	140	94
AG-0032F. .	GGCACTGGCCAGTTGGACAC	AAATTAATAAAACACATTTGAGTGTTAATAATAC	No	Yes	471	-	120	18
AG-0090F. .	AGGAATAATCGCTTTTGCCACTG	CCTCTTGATATGTCTTGCGAAATG	Sp	No	654	-	130	62
AG-0211F. .	TCGAGTTGGACGAGAGCTGTTACG	ATTTGTGCTATTCGTTTGTCACTG	No	Yes	412	-	130	81
AG-0254P. .	GTCACCGAGCTACAAAACAACACG	TCTCTTGATATGCTCGCCTTTCTC	SgP	Yes	740	147	132	94
AG-0334P. .	TGCTGCGAATATGGAGGAG	CATGGTGCTTTTCGTGCTCTT	Sp	Yes	1057	927	118	126

* Primer "3372" from Simon *et al.* 1994.

** including 27 haplotypes from Willett *et al.* 1997.

[a] Indicates if protein is likely to be secreted. Sp=present in spermatophore; SgP=signal peptide present.

[b] indicates if expression is male biased.

[c] total number of nucleotides (bp) sequenced.

[d] length of coding region (for some loci only intronic or other non-coding regions were sequenced)

[e] total number of haplotypes sequenced.

[f] number of variable sites across all samples.

Phylogenetic analyses

For the *mtDNA* locus, phylogeny reconstruction was carried out using MRBAYES version 3.1.2 (Huelsenbeck and Ronquist 2001). Searches were run for five million generations, sampling every 100 generations and discarding the first 1,000,000 generations (burn-in time). We used default priors and the general time reversible model with invariant sites and gamma rates (GTR+I+G), allowing the rate at each site to change over evolutionary history. We compared these results with those obtained using site specific rates (SSR), with sites at each codon position and in the tRNA following a gamma distribution and allowing a proportion of sites to be invariant. Because there were no differences between the two models we only show results for the SSR model. In addition to our own sequences we also included 27 sequences from Willett *et al.* (1997). The phylogenetic tree was rooted using seven *G. rubens* sequences.

We used PAUP* software package, version 4.0b10 (Swofford 2003) to reconstruct nuclear gene trees using the neighbor-joining (NJ) algorithm. To estimate distances for NJ trees for each locus, we applied hierarchical likelihood ratio tests with the program MODELTEST 3.06 (Posada and Crandall 1998). We excluded all haplotypes missing more than 10% of variable sites. Missing data were a problem for some loci due to the presence of multiple indels. Evidence of recombination (Table 4.3) complicates the interpretation of phylogenetic history, because different regions of each locus have different histories. Recombination strongly affects phylogeny reconstruction using maximum parsimony or maximum likelihood; by using a distance based approach, alleles are grouped on overall similarity and reflect approximate patterns of relatedness, although the trees may not represent a "true genealogy". We used 1000 bootstrap replicates to assess support for interior branches.

We tested the homogeneity of the phylogenetic signal among the nine nuclear loci using the partition homogeneity test (Farris *et al.* 1995). We combined data from the 47

Table 4.3 Polymorphism statistics for each loci, based on only pure populations of each species (*G. pennsylvanicus*: ITH, SCO, SCR; *G. firmus*: GUI, PAR, TRI).

Locus	Species	n[a]	L[b]	Cod[b]	S[c]	syn[d]	rep[d]	π[e]	Dxy[f]	D[g]	Rm[h]
mtDNA . ..	*firmus*	16	1825	1119	25	10	3	0.00342	0.00576	-0.6971	-
	pennsylvanicus	14	1834	1125	10	3	2	0.00119		-1.1859	-
EF-1α . .	*firmus*	20	713	369	20	3	0	0.00927	0.00871	1.1396	0
	pennsylvanicus	18	716	372	18	4	0	0.00361		-2.1637*	0
GuKc. . . .	*firmus*	30	431	153	7	2	1	0.00415	0.00410	-0.3438	1
	pennsylvanicus	24	432	156	4	1	1	0.00263		0.1699	0
Hex	*firmus*	34	460	282	10	2	2	0.00314	0.00436	-1.2695	0
	pennsylvanicus	26	460	282	2	0	0	0.00121		0.1341	0
AG-0005F	*firmus*	30	875	875	32	9	23	0.00765	0.01671	-0.6217	3
	pennsylvanicus	26	871	871	31	15	16	0.00733		-0.7963	4
AG-0032F.	*firmus*	26	444	0	4	-	-	0.00192	0.00483	-0.9905	0
	pennsylvanicus	24	457	0	9	-	-	0.00557		0.1863	2
AG-0090F	*firmus*	22	621	0	27	-	-	0.01196	0.01365	0.0113	5
	pennsylvanicus	30	570	0	22	-	-	0.01250		0.9952	8
AG-0211F	*firmus*	28	409	0	33	-	-	0.01279	0.01158	-1.5466	2
	pennsylvanicus	28	409	0	30	-	-	0.00976		-1.7644	0
AG-0254P	*firmus*	24	543	147	30	1	5	0.01218	0.01336	-0.0665	6
	pennsylvanicus	28	550	144	25	1	4	0.01245		0.0909	6
AG-0334P	*firmus*	28	1049	915	49	9	36	0.00794	0.00912	-1.3263	6
	pennsylvanicus	30	1049	909	35	8	22	0.00595		-1.0753	4

[a] Number of haplotypes analyzed
[b] Length of sequence and coding region analyzed.
[c] Total number of polymorphic sites.
[d] Number synonymous and replacement changes.
[e] Average number of nucleotide differences per site.
[f] Average number of nucleotide substitutions per site between species.
[g] Tajima's statistic (1989). *p<0.01.
[h] Minimum number of recombination events per locus.

individuals that were sequenced for all nine loci and performed the test with 1000 replicates.

Molecular population genetics

We used DNAsp v.4.20.2 (Rozas *et al.* 2003) for basic polymorphism analysis. Indels were not included in these analyses. Analyses of Molecular Variance (AMOVA, Excoffier *et al.* 1992) for pure species populations (i.e. GUI, ITH, PAR, SCO, SCR, TRI) were conducted using Arlequin v. 2.000 (Schneider *et al.* 2000). Tajima's *D* (Tajima 1989) was calculated to test for departures from neutrality with DNAsp v4.20.2. This test is based on the expectation that under mutation-drift equilibrium θ and π should be the same parameter (i.e., $4N_e\mu$). Tajima's *D* can detect signatures of recent demographic events, such as population expansion (excess of low frequency polymorphisms leading to negative Tajima's *D* values), and/or selective events (selective sweeps, negative Tajima's *D* values). By testing many loci it is possible to distinguish between the demographic and selective scenarios because a population expansion is expected to affect the entire genome, whereas selection should only affect the selected locus and adjacent (hitchhiking) regions.

Test of Selection

The relative rate of fixation of non-synonymous (d_N) and synonymous (d_S) substitutions provides an estimate of selection pressures acting on a given protein. For any set of amino acid residues, when $d_N/d_S = \omega = 1$, a neutral model of evolution cannot be rejected, whereas $\omega < 1$ indicates purifying selection, and $\omega > 1$ indicates positive selection. Although the selection parameter ω is commonly calculated using phylogenetic likelihood methods (Goldman and Yang 1994), these methods are unreliable in the presence of recombination because this process leads to not one, but multiple evolutionary trees along the gene sequence (Anisinova *et al.* 2003; Wilson and McVean 2006). In this paper we used the method recently developed by Wilson and McVean (2006) to calculate

ω in the presence of recombination. This method relaxes the assumption of a single common history for all codons, and performs Bayesian inferences of ω using a population genetics approximation to the coalescent with recombination (Hudson 1983; Li and Stephens 2003). One disadvantage of this method is that it does not provide estimates of d_N and d_S. Using OmegaMap (Wilson and McVean 2006) we estimated the selection parameter (ω), recombination rate (ρ), transition-transversion ratio (κ), and the rate of synonymous transversion μ for each gene for which we sequenced all or part of the coding region (*EF1-α*, *GuKc*, *Hex*, *AG-0005F*, *AG-0254P* and *AG-0334P*). For the other three genes, only intronic regions were sequenced (see Table 4.3). We used improper inverse prior distributions for all parameters with means $\omega = 1$, $\rho = 0.07$, $\kappa = 3.6$, $\mu = 0.3$. Both, ω and ρ were modeled as constant (*i.e.* all sites are assumed to share common values). The frequency of codons was assumed to be equal, and the number of alignment orderings was set to 10. We ran at least 250,000 iterations with a 10,000 burn-in and a thinning of 100. For each gene, two independent convergent runs were merged to provide the posterior distributions of the estimated parameters. The effective sample size for the estimated parameters was always >100, suggesting that the MCMC chains were run long enough to obtain accurate estimates. For each distribution of ω values we calculated the mode and posterior probability of selection with a cut off at $\omega>0.5$ (Swanson *et al.* 2004 showed that upon closer examination many of the loci with $\omega>0.5$ were actually under selection).

Isolation and introgression

We calculated migration rates between pure species populations (GUI, TRI and PAR for *G. firmus* and ITH, SCR and SCO for *G. pennsylvanicus*) as a proxy for gene introgression across the hybrid zone. Because genomes are mosaics and both species exclusivity and introgression are locus specific, we initially tested each locus independently. To discriminate between the relative effects of divergence and gene flow, we analyzed the loci under the isolation-with-migration analytic (IMa) model (Nielsen and

Wakeley 2001; Hey and Nielsen 2004; Hey and Nielsen 2007). This model assumes that an ancestral population splits into two populations (species) with gene flow possibly continuing between the diverging populations. To fit the IMa model to data, we used a Bayesian coalescent method that approximates the integration over the possible genealogies using a Markov chain Monte Carlo (MCMC) simulation. This method estimates marginal and posterior probability distributions for demographic parameters including directional migration rates scaled by mutation rate for the entire locus ($m1 = m_1/\mu_i$ and $m2 = m_2/\mu_i$), divergence time scaled by mutation ($t = t\mu_i$), and effective population sizes of the two species and the ancestral population (Hey and Nielsen 2004; Hey and Nielsen 2007). We obtained asymmetric estimates of migration rates between species (effective number of migrants per generation, $2N_em_i$) from the product of m_i and $\theta_i/2$.

We conducted the analysis using the Hasegawa–Kishino–Yano (HKY) model and uninformative prior distributions of parameters. To improve mixing, we used a geometric heating scheme with 50-80 parallel chains. At least 25,000 genealogies were sampled from the primary chain after a 2-5 hours burn-in. We replicated each analysis at least three times and all replicates yielded nearly identical estimates. Convergence upon the stationary distribution was assessed by estimating the effective sample size (ESS) and autocorrelation of parameter values measured over the course of the run. The analysis was considered to have converged upon a stationary distribution if the independent runs generated similar posterior distributions (Hey 2005), with a minimum ESS of 100 (Kuhner and Smith 2007). For credibility intervals, we report the 90% highest posterior density (HPD) interval, the interval which includes 90% of the probability density of a parameter.

To test for differences between *m*1 and *m*2 in nuclear genes we used a likelihood ratio test of nested models (L mode option in IMa). We separated loci into those not under selection and those likely under selection (*AG-0005F* and *AG-0334P* – see Results).

We ran two analyses, one combining all nuclear loci except for *AG-0005F* and *AG-0334P* and the other combining the two loci under selection (i.e. *AG-0005F* and *AG-0334P*). Our aim was to test if introgression rates for neutral loci vary in directionality and to see how introgression rates of loci encoding reproductive proteins under selection differ from those of neutral loci. Although IMa assumes no selection, our intention was not to calculate actual migration rates but to compare relative introgression rates between different sets of genes. Here we use migration as an approximation to introgression rates between parts of the genome subjected to different evolutionary pressures.

Because the IMa model assumes no recombination we used only the longest non-recombining region for each locus. Non recombining regions were inferred using the algorithms implemented in the IMgc software package (Woerner *et al.* 2007).

RESULTS

Phylogenetic analyses

For the *mtDNA* data, the Bayesian posterior probability approach, using the GTR+I+G model or the SSR model, produced a tree with six major haplotype groups (Figure 4.2). These six clades correspond to the haplotype groups identified by Willett *et al.* (1997) except that one of the previously identified groups (northern *G. firmus*) has split into three. We thus use the nomenclature of Willett *et al.* (1997), combining the three basal clades into a northern *G. firmus* group (Figure 4.2). We refer to groups as (1) Northern *G. pennsylvanicus* (including ITH, SCO, SCR), (2) southern *G. pennsylvanicus* (including COR and ASH from Willett *et al.* 1997), (3) northern *G. firmus* (including GUI, and TRI) and (4) southern *G. firmus* (including PAN from Willett *et al.* 1997 and PAR). In both tree figures (Figure 4.2 and Figure 4.3) we use colored ovals to designate individuals from apparently pure species populations (blue and yellow for *G. firmus* and *G. pennsylvanicus,* respectively) and open squares to represent mixed populations. In the

Figure 4.2 Posterior probability tree of mtDNA, partitioned by site specific rates (SSR). The search with the program MrBayes (Huelsenbeck & Ronquist 2001) was run for five million generations, discarding the first one million generations. We used default priors, a GTR model, invariant sites and gamma rates. Haplotypes in italics are from Willett *et al.* (1997). Yellow and red represent northern and southern *G. pennsylvanicus* clades and blue and green represent northern and southern *G. firmus* clades. Yellow and blue ovals represent individuals from pure populations of each species and squares represent mixed hybrid populations.

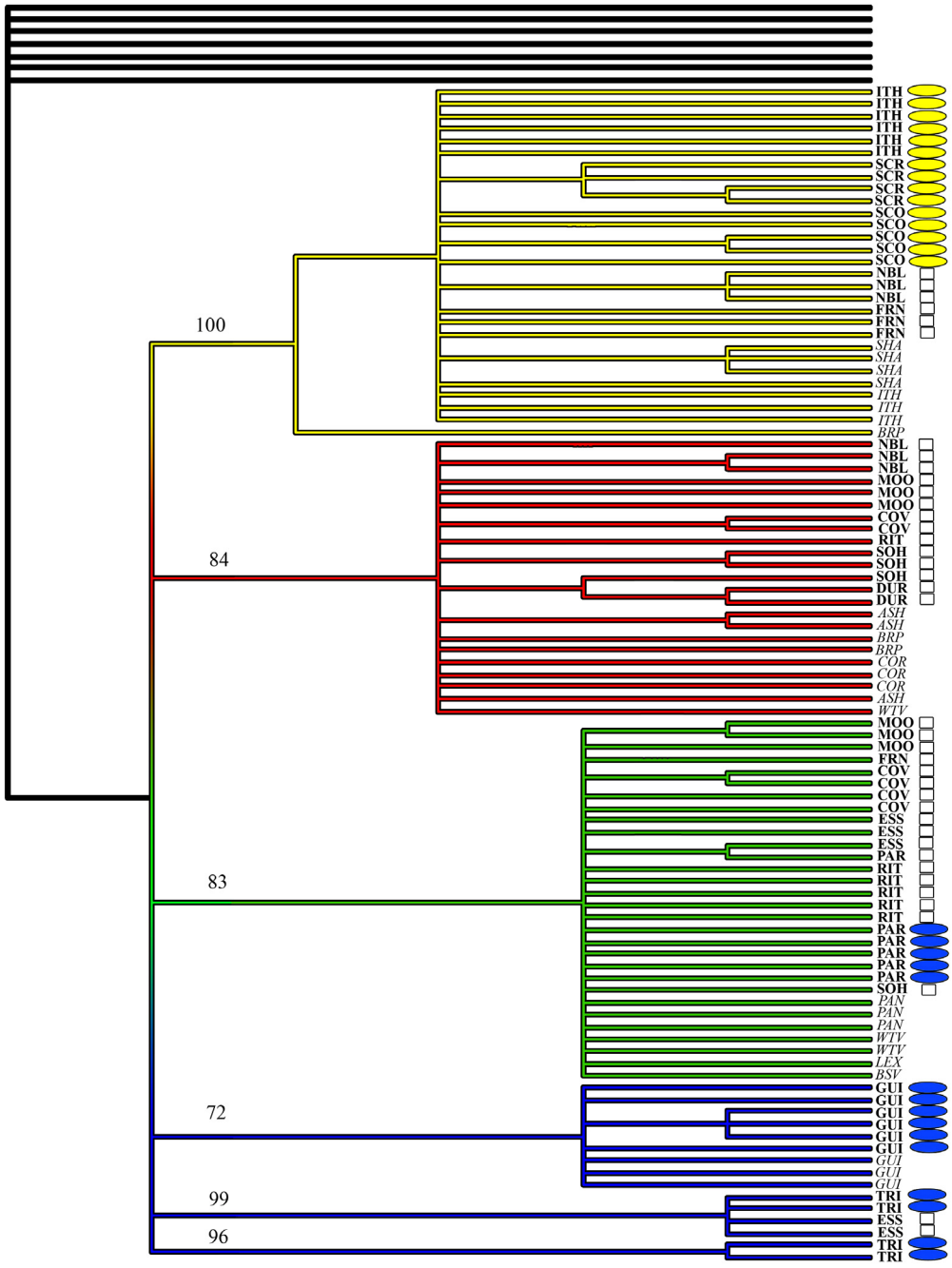
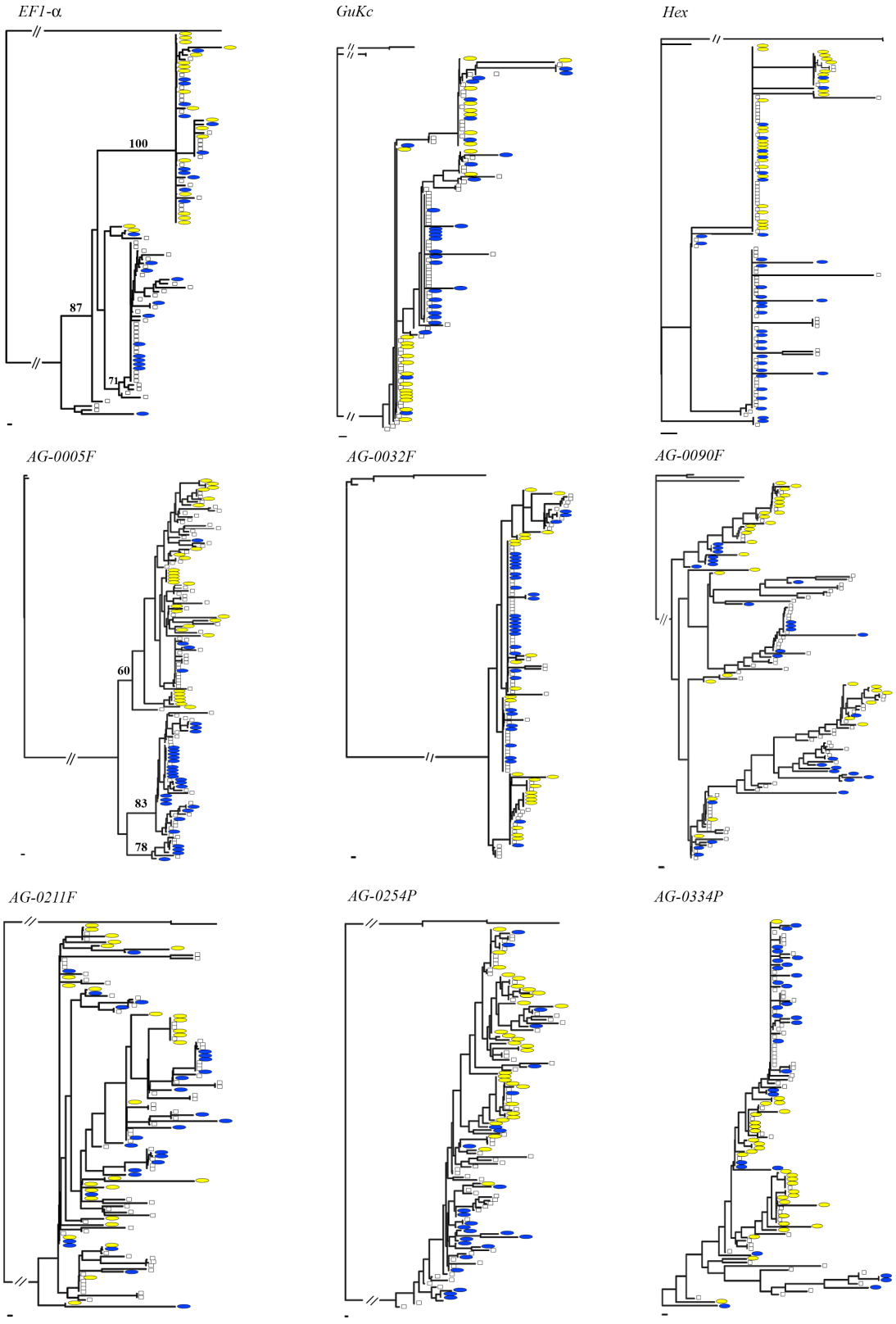

100
84
83
72
99
96
ITH
ITH
ITH
ITH
ITH
ITH
SCR
SCR
SCR
SCR
SCO
SCO
SCO
SCO
SCO
NBL
NBL
NBL
FRN
FRN
FRN
SHA
SHA
SHA
SHA
ITH
ITH
ITH
BRP
NBL
NBL
NBL
MOO
MOO
MOO
COV
COV
RIT
SOH
SOH
SOH
DUR
DUR
ASH
ASH
BRP
BRP
COR
COR
COR
ASH
WTV
MOO
MOO
MOO
FRN
COV
COV
COV
COV
ESS
ESS
ESS
PAR
RIT
RIT
RIT
RIT
RIT
PAR
PAR
PAR
PAR
PAR
SOH
PAN
PAN
PAN
WTV
WTV
LEX
BSV
GUI
GUI
GUI
GUI
GUI
GUI
GUI
GUI
GUI
TRI
TRI
ESS
ESS
TRI
TRI

Figure 4.3 Neighbor Joining trees for all nuclear loci. Distances were calculated with the locus specific model selected by MODELTEST 3.06 (see Table 4.4). Ovals represent pure *G. firmus* (blue) and *G. pennsylvanicus* (yellow) populations. Open squares represent mixed populations. Bars at the bottom of each tree represent 0.0005 substitutions per site. Most loci are rooted only with *G. rubens*. *EF1*-α, *GuKc* and *Hex* are rooted with *G. rubens* and *G. bimaculatus*. AG-0334P is unrooted; in the rooted version haplotypes are collapsed due to the huge branch leading to *G. bimaculatus* (see Figure 4.12). Figures 4.4 – 4.12 show trees labeled with population names.


EF1-α
100
87
71
GuKc
Hex
AG-0005F
60
83
78
AG-0032F
AG-0090F
AG-0211F
AG-0254P
AG-0334P

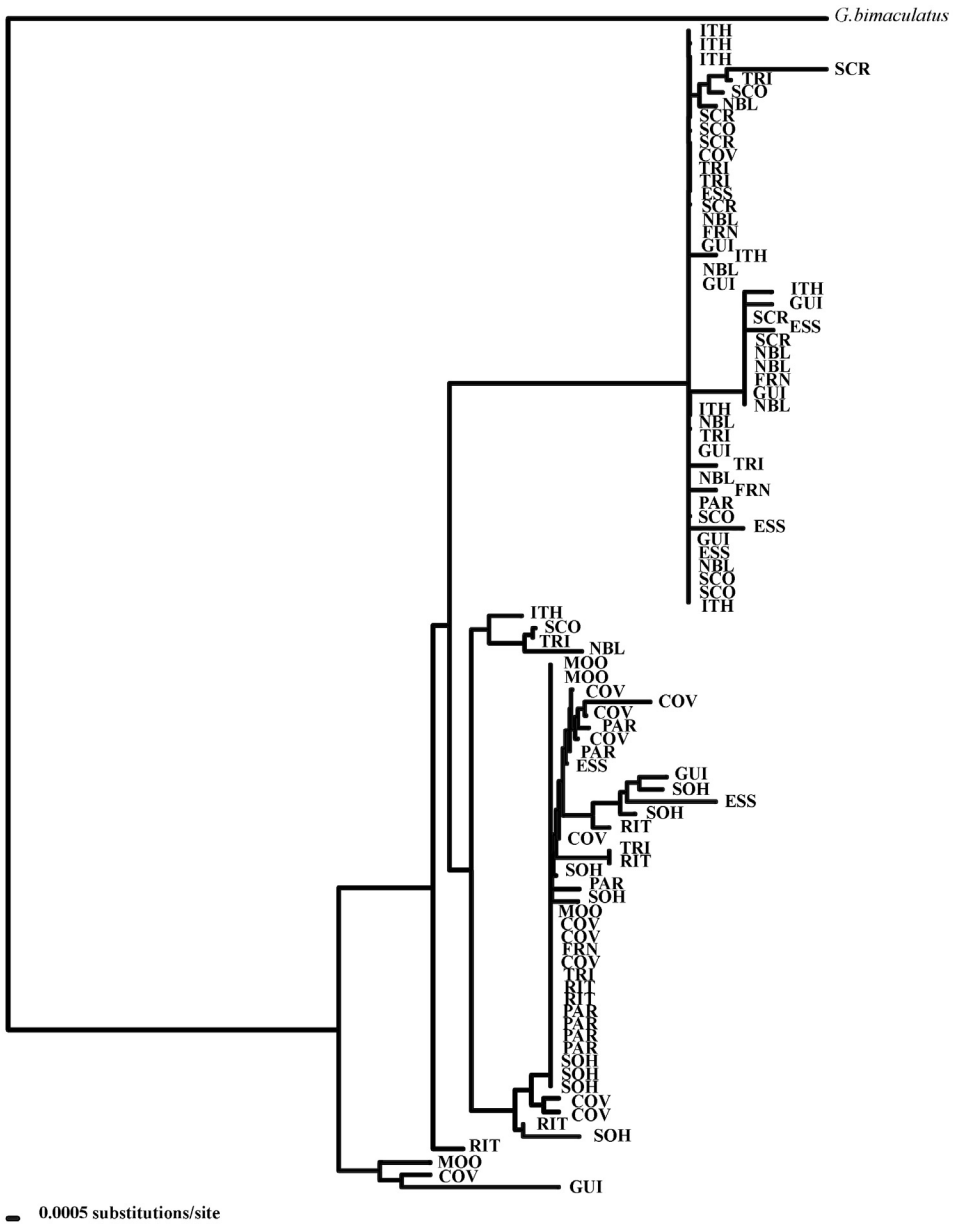


Figure 4.4 *EF1*-α Neighbor Joining tree. Distances were calculated with *EF1*-α specific model selected by MODELTEST 3.06 (see Table 4.4). All population names shown (for abbreviation see Table 4.1).

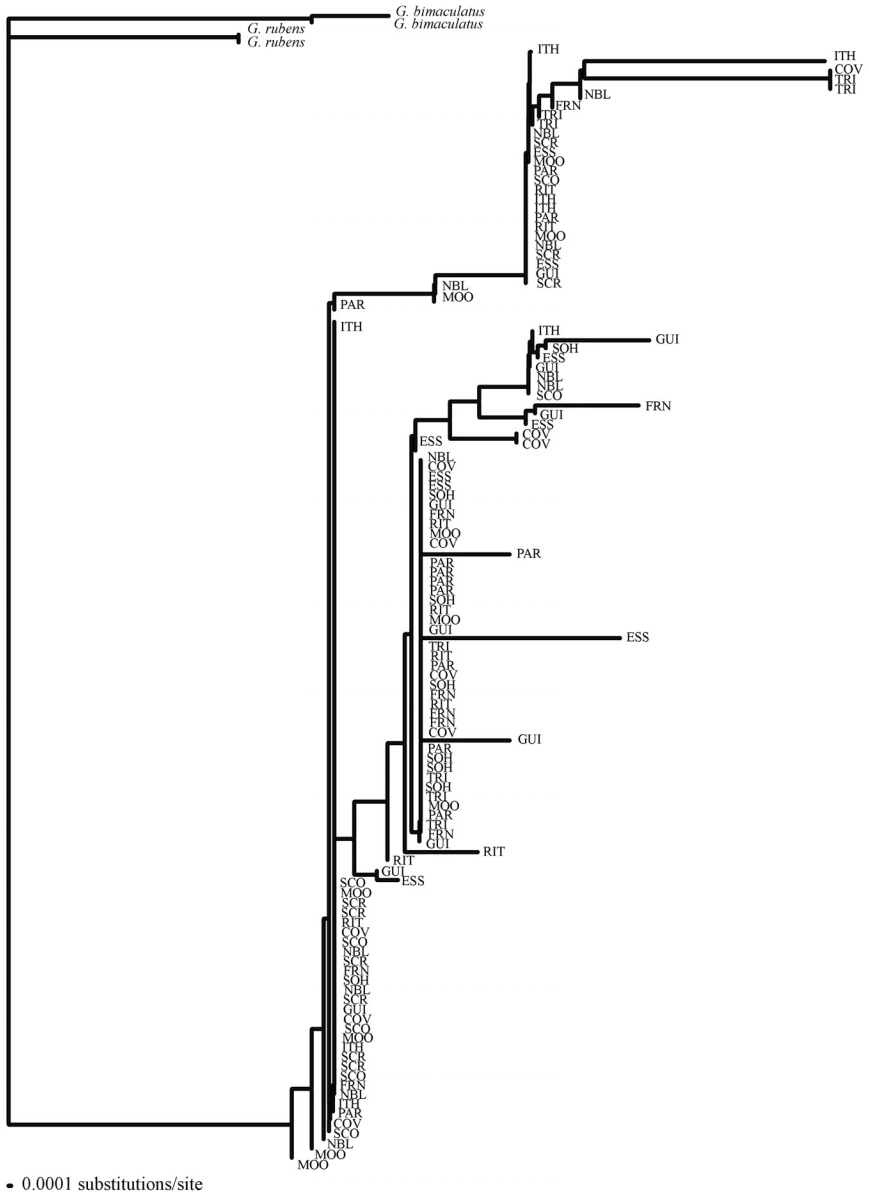


Figure 4.5 *GuKc* Neighbor Joining tree. Distances were calculated with *GuKc* specific model selected by MODELTEST 3.06 (see Table 4.4). All population names shown (for abbreviation see Table 4.1).

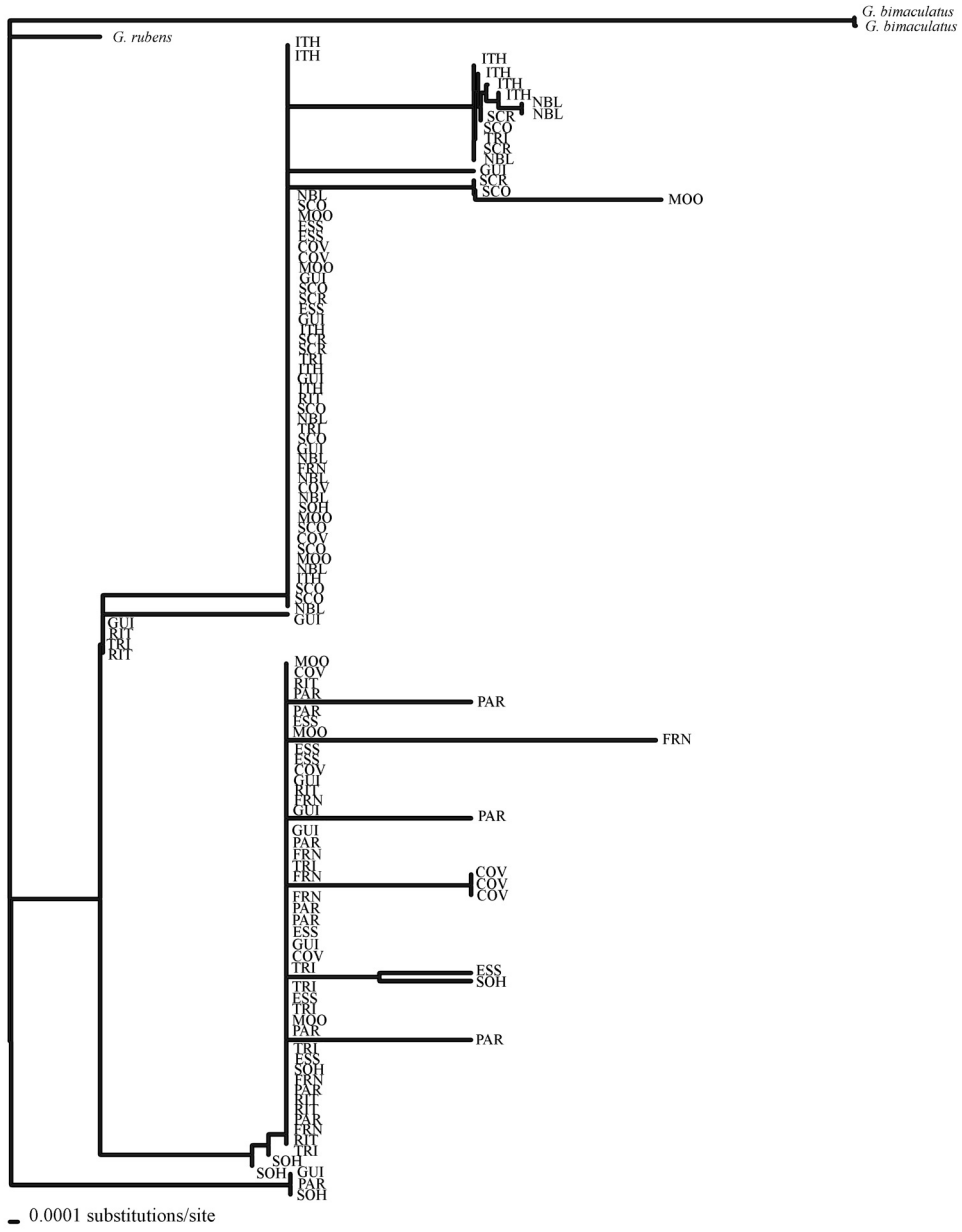


Figure 4.6 *Hex* Neighbor Joining tree. Distances were calculated with *Hex* specific model selected by MODELTEST 3.06 (see Table 4.4). All population names shown (for abbreviation see Table 4.1).

Figure 4.7 *AG-0005F* Neighbor Joining tree. Distances were calculated with *AG-0005F* specific model selected by MODELTEST 3.06 (see Table 4.4). All population names shown (for abbreviation see Table 4.1).

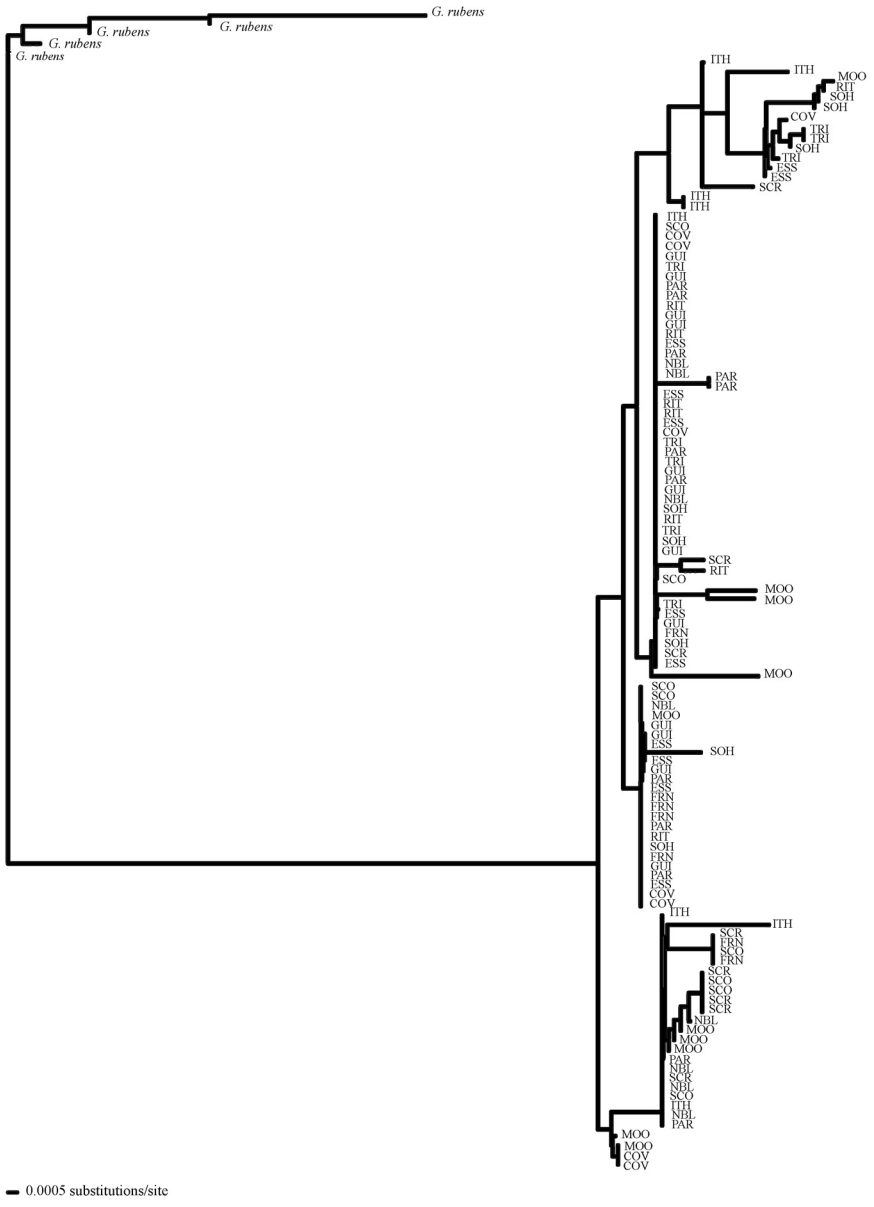


Figure 4.8 *AG-0032F* Neighbor Joining tree. Distances were calculated with *AG-0032F* specific model selected by MODELTEST 3.06 (see Table 4.4). All population names shown (for abbreviation see Table 4.1).

Figure 4.9 *AG-0090F* Neighbor Joining tree. Distances were calculated with *AG-0090F* specific model selected by MODELTEST 3.06 (see Table 4.4). All population names shown (for abbreviation see Table 4.1).

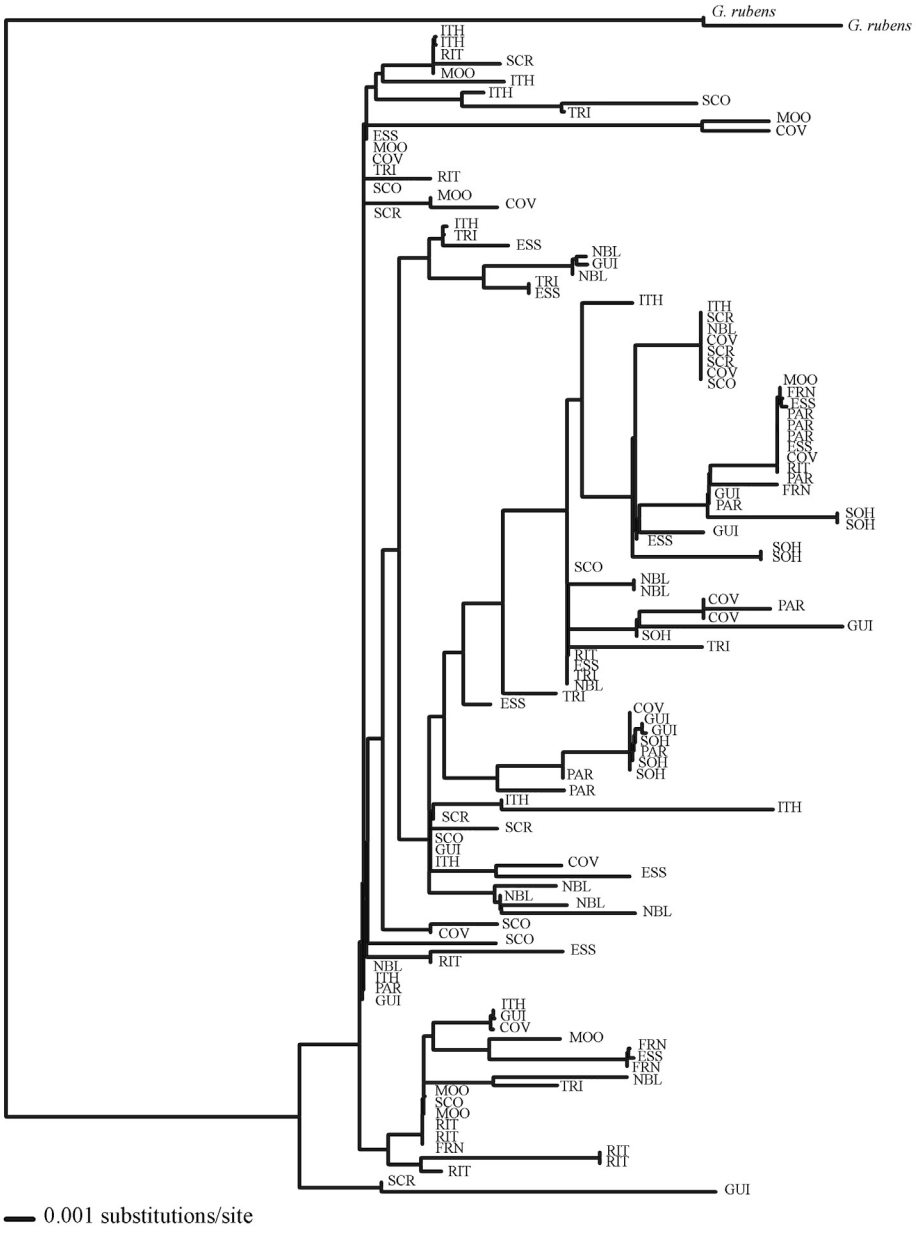


Figure 4.10 *AG-0211F* Neighbor Joining tree. Distances were calculated with *AG-0211F* specific model selected by MODELTEST 3.06 (see Table 4.4). All population names shown (for abbreviation see Table 4.1).

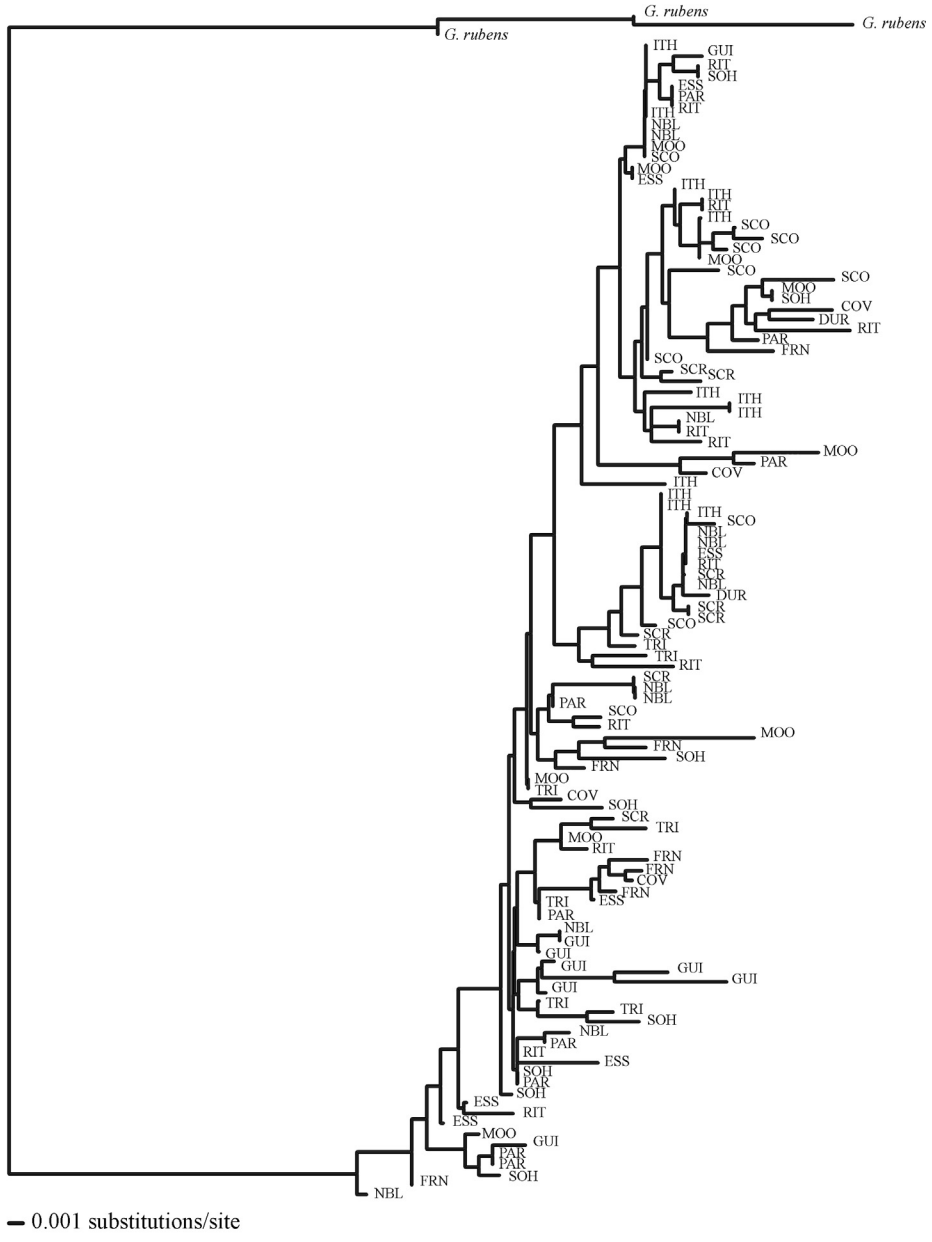


Figure 4.11 *AG-0254P* Neighbor Joining tree. Distances were calculated with *AG-0254P* specific model selected by MODELTEST 3.06 (see Table 4.4). All population names shown (for abbreviation see Table 4.1).

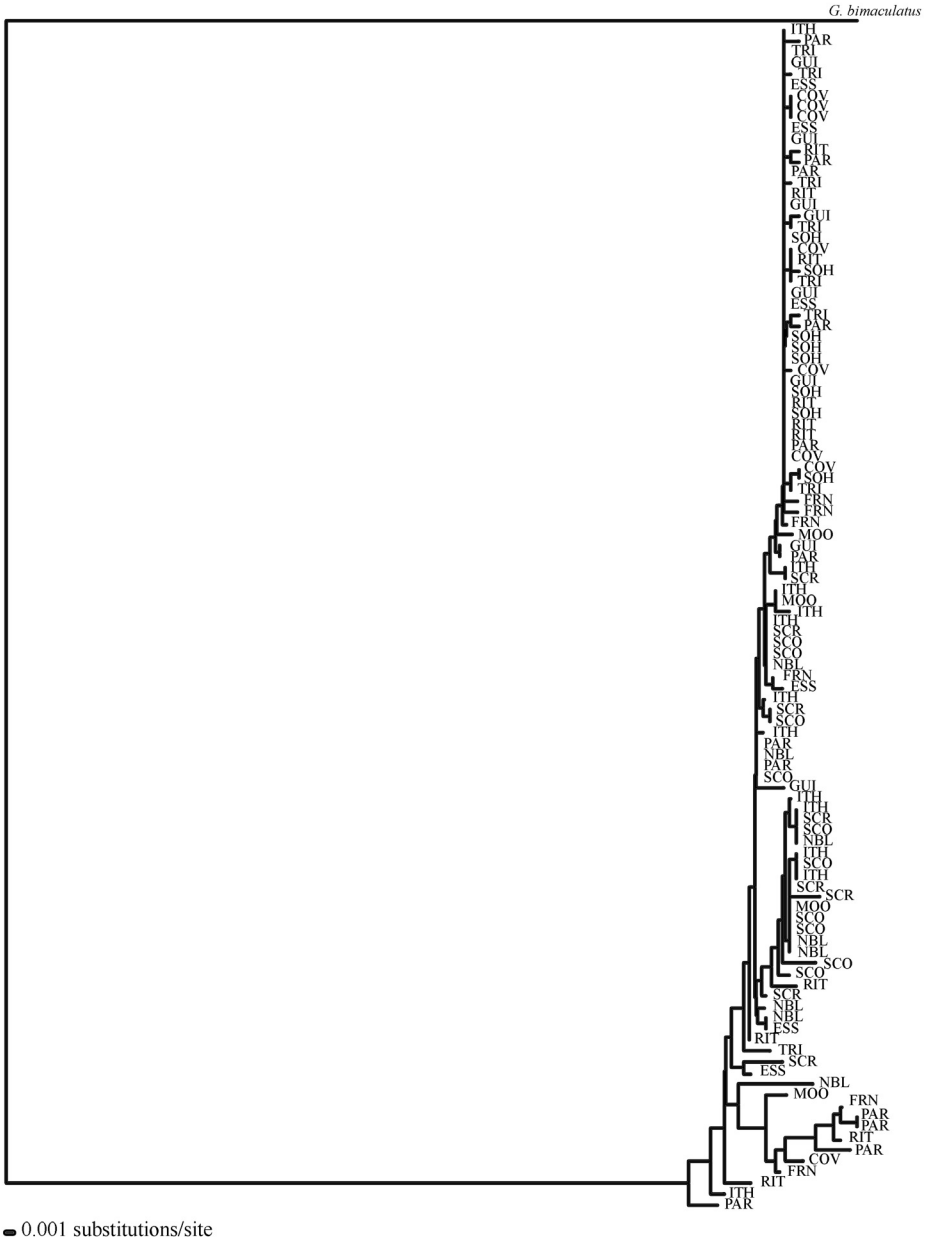


Figure 4.12 *AG-0334P* Neighbor Joining tree. Distances were calculated with *AG-0334P* specific model selected by MODELTEST 3.06 (see Table 4.4). All population names shown (for abbreviation see Table 4.1).

mtDNA phylogeny we color each of the four major groups (green and blue for *G. firmus* and red and yellow for *G. pennsylvanicus*) and use these colors in Figure 4.1 to represent the percentage of crickets belonging to each *mtDNA* clade in each of the populations.

For nuclear loci the most complex evolution model selected by MODELTEST 3.06 (Posada and Crandall 1998) was the Tamura-Nei (Tamura and Nei 1993) with Gamma rates (Table 4.4). The null hypothesis of homogeneity of the phylogenetic signal among nuclear loci was rejected (partition-homogeneity test, $p<0.001$). All nuclear gene trees are shown in Figure 4.3. More detailed versions of all trees, including population information, can be found in supplementary materials (Figures. 4.4 – 4.12). We excluded positions 294-463 from *AG-0254P,* which were missing data in a substantial fraction of the haplotypes (n=44) due to two indels located within an intron.

Only *EF1-α* and *AG-0005F* had strong bootstrap support (>70%) for major branches. *EF1-α* had two major clades with bootstrap support higher than 70% (Figure 4.3). These clades are largely defined by an intron polymorphism with two alternative alleles. One clade includes mostly haplotypes from northern populations (GUI, ITH, NBL, SCO, SCR and TRI) and the other includes mostly haplotypes from southern populations (COV, ESS, FRN, MOO, PAR, RIT and SOH). In the predominantly northern group, only nine haplotypes (out of 47) are from southern populations (COV, ESS and FRN). In the predominantly southern group, only two haplotypes (out of 39) are from northern populations (GUI and TRI). A small basal clade includes a mix of southern and northern population haplotypes (COV, GUI and MOO). Because all of the pure *G. pennsylvanicus* populations that we sampled are in the north of the species range (Figure 4.1), the predominantly northern group has an overrepresentation of *G. pennsylvanicus* haplotypes.

The genealogy for *AG-0005F* reveals a clear separation of pure species *G. firmus* and *G. pennsylvanicus* haplotypes. Two sister clades with bootstrap support higher than

Table 4.4 Selected model for each loci (using MODELTEST 3.06)

Loci	model	gamma[a]	T/V[b]	I[c]	A[d]	C[d]	G[d]
EF1-α.....	HKY + G	0.5680	0.8613	0	0.2973	0.1475	0.2142
GuKc.....	F81 + G	0.0163	1	0	0.2762	0.1761	0.2227
Hex….....	F81	-	1	0	0.3262	0.1519	0.1894
AG-0005F	HKY + I + G	0.7978	1.2434	0.7412	0.2484	0.3398	0.2162
AG-0032F	F81 + I + G	0.7058	1	0.7767	0.3022	0.2081	0.1633
AG-0090F	F81 + I + G	0.7203	1	0.8113	0.3406	0.1688	0.1781
AG-0211F	F81 + G	0.8229	1	0	0.3194	0.1933	0.1728
AG-0254P	HKY + I + G	0.6329	0.9484	0.7026	0.2566	0.2041	0.2365
AG-0334P	TrN + G	0.3078	*	0	0.3098	0.1625	0.2265

[a] Gamma distribution shape

[b] Transition/transversion rate

[c] Proportion of invariable sites

[d] Proportion of bases.

* A/C=1.0000, A/G=1.0822, A/T= 1.0000 C/G=1.0000 C/T=2.5831

70% (Figure 4.3d) only include haplotypes from mixed populations and from pure *G. firmus* populations. All haplotypes from pure *G. pennsylvanicus* populations belong to a third major clade with lower bootstrap support (60%), which also includes four haplotypes (out of 78) from pure *G. firmus* populations. All mixed or hybrid populations, except for NBL, have haplotypes in both the predominantly *G. pennsylvanicus* clade and in the *G. firmus* clades. All six haplotypes from NBL are in the predominantly *G. pennsylvanicus* clade.

Molecular population genetics

Polymorphism analyses for *G. firmus* and *G. pennsylvanicus,* using only pure populations (*G. firmus*: GUI, PAR and TRI; *G. pennsylvanicus*: ITH, SCO, SCR), are summarized in Table 4.3. In general, *G. firmus* has more nucleotide variation suggesting larger population sizes and perhaps more ancient populations (see Discussion). The most noteworthy observation for nuclear genes is the high number of replacement substitutions for *AG-0005F* and *AG-0334P*. Tajima's *D* was not significant for any locus/species or locus/ population combination except *EF1-α/G. pennsylvanicus*, which had a significant negative value (Table 4.3). Signs of demographic expansion were not evident as there were no consistent trends towards positive or negative Tajima's *D* values across loci.

For most nuclear genes the average nucleotide differences per site (π) for at least one of the species was equal to or greater than the average differences per site between species (*D*xy, Table 4.5). Only *Hex, AG-0005F* and *AG-0334P* had substantially higher *D*xy values than π values. Surprisingly there were no diagnostic sites at the species level for any of the nuclear loci. For *mtDNA* there are fixed differences between clades within and between species.

AMOVA analyses (Excoffier *et al.* 1992) showed that almost all variation is due to within population variation (all F_{ST} covariance components are significant). Even for *mtDNA*, only 14% of the variation can be attributed to differences between species and the

Table 4.5 ANOVA and Hierarchical Analyses for pure populations of *G. firmus* and *G. pennsylvanicus*

Source of variation (%)	*mtDNA*	*EF1-α*	*GuKc*	*Hex*	*AG-0005F*	*AG-0032F*	*AG-0090F*	*AG-0211F*	*AG-0254P*	*AG-0334P*
Among species……..	14.13	11.94	16.16	47.74	43.20	18.56	5.33	3.93	9.99	22.39
Among populations within species	8.77	14.78	3.75	5.60	6.04	1.62	7.09	4.24	6.10	3.66
Within populations……	77.10	73.28	80.09	46.66	50.76	79.82	87.58	91.83	83.91	73.95
Fixation indices										
F_{CT} (species/total)…….	0.141	0.119	0.162	0.477	0.432	0.186	0.053	0.039	0.100	0.224
F_{SC} (population/species)……	0.102	0.168*	0.045	0.107*	0.106**	0.020	0.075*	0.044	0.068*	0.047
F_{ST} (population/total)	0.229*	0.267**	0.199**	0.533***	0.492**	0.202***	0.124**	0.082*	0.161***	0.261***

* P<0.05

** P<0.01

*** P<0.001

covariance component is not significant (F_{CT}). For nuclear loci, only *Hex*, *AG-0005F*, and *AG-0334P* have more than 20% of the total variation attributed to between species variation; however none of the between species covariance components (F_{CT}) are significant. In general there is very little population structure within species, most loci have less than 10% of the total variation attributable to among populations variation. *EF1-α* is an exception to this pattern with 15% of the total variation attributed to among populations within species variation. This pattern seems to be caused by two very different alleles with very different proportions in northern and southern populations.

Tests of selection

To estimate selection on amino acid sequence, we used sequences from all populations, but could only carry out the test for the six nuclear loci for which sequences from coding regions were available. The numbers of coding (and variable) sites analyzed for each locus were 372 (13) for *EF1-α,* 156 (5) for *GuKc,* 282 (7) for *Hex,* 816 (80) for *AG-0005F*, 147 (8) for *AG-0254P* and 924 (78) for *AG-0334P*.

We calculated the d_N/d_S ratio using the Nei and Gojobori (1986) equation as implemented in DNAsp v.4.20.2 (Rozas *et al.* 2003). The d_N/d_S ratios for the six genes were: *EF1-α* =0.02, *GuKc* = 0.24, *Hex*= 0.36, *AG-0005F*= 0.65, *AG-0254P* =0.55 and *AG-0334P* =1.21. Swanson *et al.* (2004) showed that statistical evidence for adaptive evolution at some codons can be found for most genes having overall gene $d_N/d_S > 0.5$. Of our loci only *AG-0334P* and *AG-0005F* had d_N/d_S ratios substantially higher than 0.5 and are thus candidates to be under selection. However these d_N/d_S ratios are probably inaccurate since the Nei and Gojobori (1986) d_N/d_S calculation does not take into account recombination and nuclear loci have likely experienced recombination. Recombination can cause a high number of false positives in d_N/d_S ratios (Anisimova *et al.* 2003; Shriner *et al.* 2003), because trees from recombining sequences will have longer terminal branches and smaller time to the most recent common ancestor (Schierup and Hein 2000).

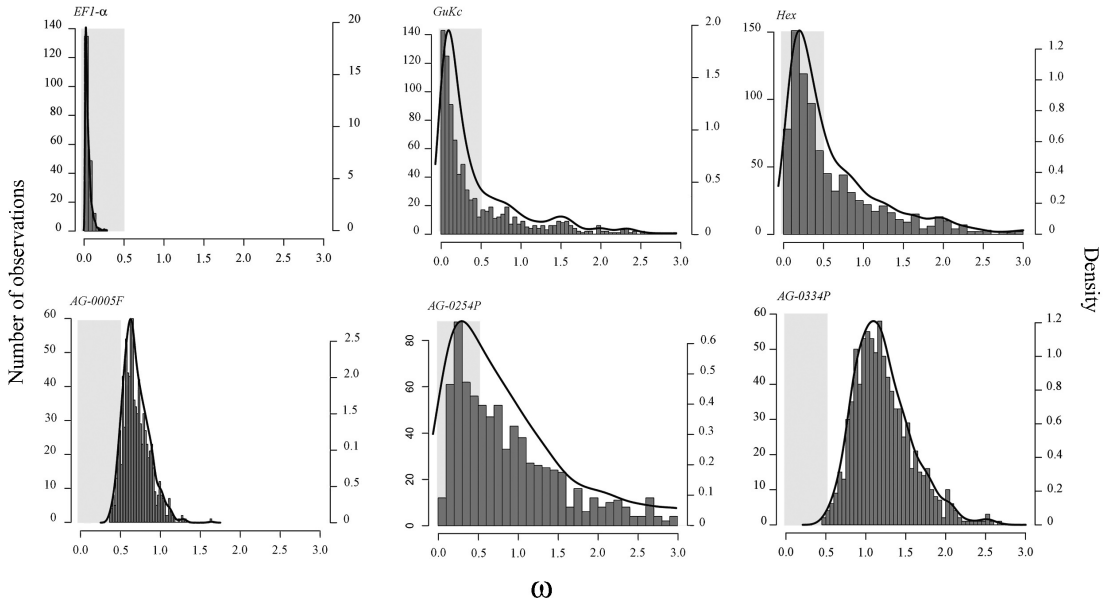


Figure 4.13 Posterior probability estimates of ω for each locus. Gray shading shows ω values below 0.5 which are unlikely to indicate selection. All loci except for *AG-0005F* and *AG-0334P* have point estimates of ω below 0.5 (see text).

To test for selection accounting for recombination, we estimated ω with the program omegaMap (Wilson and McVean 2006). Because ω distributions are not normal (Figure 4.13), here we report the mode for each locus, which, in this case, is more representative of a maximum likelihood estimate. For *EF1-α* mode =0.02, for *GuKc* mode=0.11, for *Hex* mode=0.24, for *AG-0005F* mode=0.65, for *AG-0254P* mode=0.43, and for *AG-0334P* mode=1.04. Again the only loci with ω > 0.5 are *AG-0005F* and *AG-0334P*. The probability of selection was greater than 90% only for *AG-0005F* and *AG-0334P* (0.91 and 1.00 respectively). The probabilities of selection for the other loci were zero for *EF1-α,* 0.36 for *Hex,* 0.20 for *GuKc,* and 0.69 for *AG-0254P*.

Isolation and introgression

We calculated directional migration rates between pure *G firmus* and *G. pennsylvanicus* populations as a proxy for gene introgression across the hybrid zone. We selected only non-recombining regions of each gene using at least 17 haplotypes per species (average of 25 for *G. firmus* and 23 for *G. pennsylvanicus*). The number of sites (and variable sites) analyzed for each locus were 1767 (40) for *mtDNA*, 536 (33) for *EF1-α,* 319 (9) for *GuKc,* 460 (11) for *Hex,* 578 (24) for *AG-0005F*, 443 (8) for *AG-0032F*, 390 (18) for *AG-0090F*, 320 (39) for *AG-0211F*, 284 (23) for *AG-0254P* and 797 (40) for *AG-0334P*.

Directional migration rates *m*1 and *m*2 for each locus were calculated with the isolation-with-migration analytic model (IMa) (Nielsen and Wakeley 2001; Hey and Nielsen 2004; Hey and Nielsen 2007) (see Figure 4.14 and Table 4.6). The directional migration rate *m*1 represents migration forward in time from *G. firmus* to *G. pennsylvanicus* and *m*2 represents migration forward in time from *G. pennsylvanicus* to *G. firmus*. One of the most striking patterns is the difference in migration rates across loci. For *mtDNA* there was effectively no migration in either direction. This was expected because *mtDNA* has smaller effective populations size (and thus signatures of ancestral

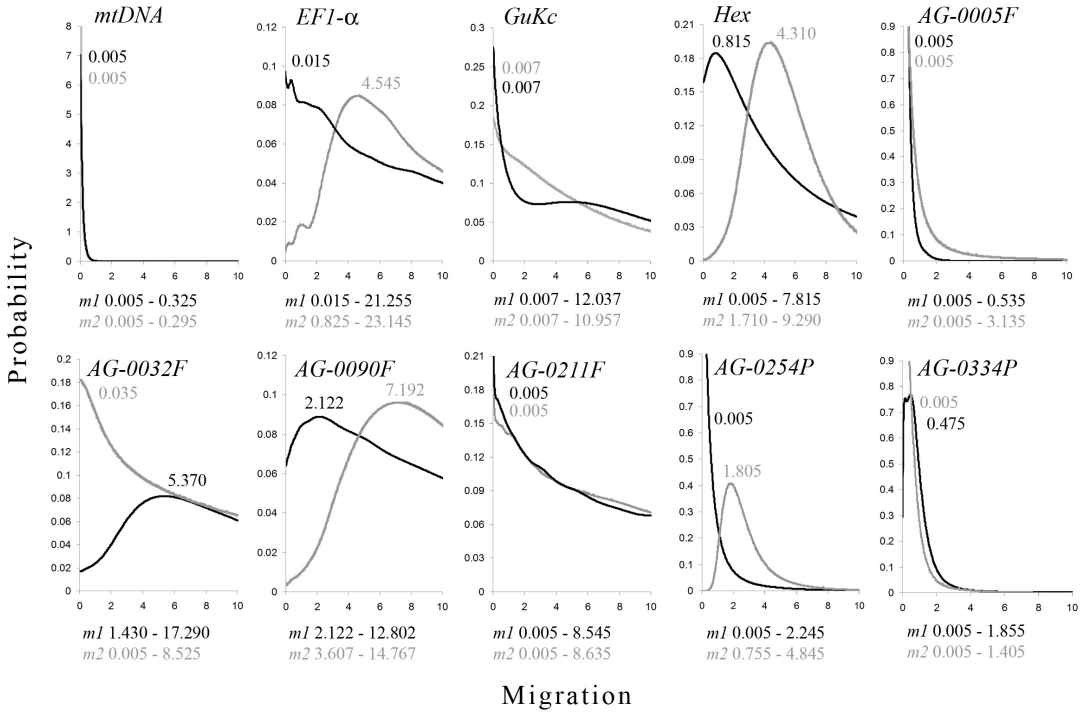


Figure 4.14 Posterior probability estimates of migration parameters (scaled by mutation rate) between *G. firmus* and *G. pennsylvanicus*. Black line shows *m1* (migration forward in time from *G. firmus* to *G. pennsylvanicus*) and grey line shows reverse migration (*m2*) for each of the analyzed loci (in mtDNA the lines are superimposed). Numbers above lines indicates the maximum likelihood value. Numbers bellow each graph show the 90% Highest Posterior Density (HPD) intervals of migration rates *m1* and *m2*.

Table 4.6 Effective migration rates and theta for *G firmus* and *G. pennsylvanicus*.

Loci	$2N_1m_1$ [a]	$2N_2m_2$ [b]	θ_1 [c]	θ_2 [d]
mtDNA. . . .	0 (0 - 2.712)	0 (0 – 6.810)	16.69	46.17
EF1-α.	0.007 (0.007- 9.235)	292.406 (53.077 – 1489.049)	0.869	128.67
GuKc.	0.003 (0.003- 4.271)	0.001 (0.001 - 1.864)	0.710	0.340
Hex.	0.048 (0 – 0.461)	29.33 (12.351 – 63.218)	0.118	13.610
AG-0005F. .	0 (0- 1.285)	0 (0 – 0.907)	4.805	0.579
AG-0032F. .	5.205 (1.386 – 16.759)	0.005 (0 – 1.901)	1.939	0.446
AG-0090F. .	0.718 (0.718- 4.334)	21.909 (10.987 – 44.982)	0.677	6.092
AG-0211F. .	0 (0 – 57.354)	0 (0 – 35.256)	13.424	8.166
AG-0254P. .	0 (0 – 1.876)	47.186 (19 737 – 126.658)	1.671	52.834
AG-0334P. .	1.197 (0 – 4.667)	0 (0 – 3.217)	5.042	4.580
Selected loci	0.085 (0 – 1.229)	0.331 (0 – 0.932)	4.869	2.323
Neutral loci	0 (0 – 0.775)	4.077 (2.111 – 6.563)	2.212	3.855

[a] Effective rate at which genes come into *G. pennsylvanicus*, per generation.

[b] Effective rate at which genes come into *G. firmus*, per generation

[c] Estimate of θ (4Nμ) for *G. pennsylvanicus*.

[d] Estimate of θ (4Nμ) for *G. firmus*.

introgression will be lost in a shorter time), is more likely to introgress from *G. pennsylvanicus* into *G. firmus* (given the direction of the incompatibility) and all tested populations are located far from the hybrid zone. The two loci likely under selection, *AG-0005F* and *AG-0334P,* also had near zero migration rates with narrow 90% HPD (Figure 4.14 and Table 4.6). For most other nuclear loci *m*2, the migration from *G. pennsylvanicus* into *G. firmus*, was higher than *m*1.

We used the nested model likelihood ratio statistics (Hey and Nielsen 2007) to test for differences between *m*1 and *m*2. To do this we combined all nuclear loci excluding *AG-0005F* and *AG-0334P* (Figure 4.15). We also tested migration rates using only *AG-0005F* and *AG-0334P* (Figure 4.15). As stated earlier, our intention was to use migration rates as a proxy for introgression across the hybrid zone. Thus, even though IMa assumes no selection, testing migration rates for the "selected" loci can give us an idea of the impact of selection on introgression rates across species boundaries. For neutral nuclear loci the model with identical migration rates $m1=m2$ was significantly rejected ($-2\Lambda=15.39$, d.f.=1, $P<0.001$), implying that *m*2 is actually higher than *m*1. For the loci under selection there was no significant difference between migration rates ($-2\Lambda=3.82$, d.f.=1, $P>0.05$) and their migration maximum likelihood estimates were very close to zero (Figure 4.15).

To get an estimate of effective population sizes we calculated θ for the neutral loci data using IMa. Using a rough mutation estimate for nuclear loci of 10^{-9} per site/generation, the estimated effective population sizes were huge; 2.6 million for *G. firmus* ($\theta = 3.85$) and 1.5 million for *G. pennsylvanicus* ($\theta = 2.25$). To get an estimate of time since divergence we used the mtDNA data, assumed 1.2 % divergence per million years per lineage (Brower 1994), and calculated time since split (*t*) with IMa. The estimated time since divergence was 202,320 years assuming one generation per year ($t/\mu = 4.29$).

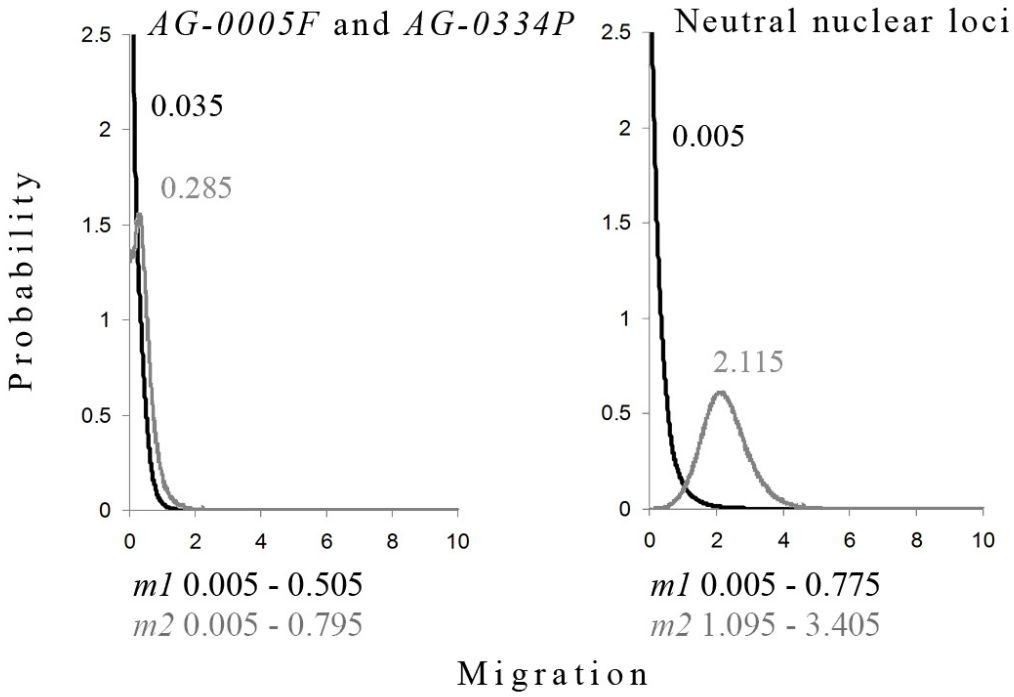


Figure 4.15 Joint posterior probability estimates of migration parameters (scaled by mutation rate) between *G. firmus* and *G. pennsylvanicus* for the two loci under selection (*AG-0005F* and *AG-0334P*) and for all other nuclear loci combined. Black line shows *m1* (migration forward in time from *G. firmus* to *G. pennsylvanicus*) and grey line shows reverse migration (*m2*) for each of the analyzed loci. Numbers above lines indicates the maximum likelihood value. Numbers below each graph show the 90% Highest Posterior Density (HPD) intervals of migration rates *m1* and *m2*.

DISCUSSION

Individuals within a species are thought to share defining properties that are not easily disturbed by hybridization and gene introgression (Coyne and Orr 2004; Templeton 1994). However, random sorting of ancestral polymorphism and differential introgression will cause recently diverged species to be mosaics with respect to molecular genealogies (Ting *et al.* 2000). These species will share alleles throughout much of their genomes. The apparent conflict between a unique species identity and widespread allele sharing disappears when we consider speciation models in which relatively few loci are responsible for the barriers to gene exchange and for species divergence. Because so-called "speciation genes" or "barrier genes" may often experience strong natural selection and are unable to cross species boundaries, they will become fixed or almost fixed in each species. It is thus expected that, across the genome of closely related species, genes will show different patterns of variation depending on their contribution to reproductive barriers, the nature of selection, and linkage relationships and recombination rates.

Introgression and selection in nuclear loci

We estimated directional migration rates between pure *G. firmus* and *G. pennsylvanicus* populations as a proxy for gene introgression across the hybrid zone. Introgression rates between the two species for the nine nuclear loci and one mitochondrial locus are strikingly different (Figure 4.14). Among the nuclear loci, only *AG-0005F* and *AG-0334P* have near zero introgression estimates with narrow 90% highest posterior densities (Figure 4.14). These two loci are also the only ones with ω values substantially greater than 0.5 and probability of selection greater than 90% (see Test of selection Results). Empirical evidence suggests that when $\omega > 0.5$ across all amino acid residues in a protein, there is a strong likelihood that selection is operating on some subset of these residues (Swanson *et al.* 2004). The joint introgression estimate for the two loci under selection is markedly different from the joint introgression estimate for

the apparently neutral loci (Figure 4.15). Although the two directional introgression rates for the selected loci do not differ (*m1* and *m2* are both near zero), directional introgression rates for neutral loci are substantially different ($P<0.0001$), with *m2,* the introgression rate forward in time from *G. pennsylvanicus* to *G. firmus,* significantly greater than *m1*. Thus, *G. pennsylvanicus* alleles are flowing into *G. firmus*, but gene flow in the other direction is significantly lower.

Differential and asymmetric introgression between *G. firmus* and *G. pennsylvanicus* has been reported for mtDNA (Harrison *et al.* 1987; Harrison and Bogdanowicz 1997; Willett *et al.* 1997) and allozymes (Harrison and Arnold 1982). Furthermore, in a fine-scale study of the hybrid zone in Connecticut, Ross and Harrison (2002) also observed differential introgression at nuclear loci, with alleles moving from *G. pennsylvanicus* into *G. firmus*. In light of recent behavioral studies (Maroja et al unpublished data), asymmetric introgression is expected; not only are hybrid offspring only produced by *G. pennsylvanicus* females, but these F1 hybrids appear to prefer to backcross to *G. firmus*. This hybrid mate choice behavior will obviously limit *m1,* the introgression rate from *G. firmus* to *G. pennsylvanicus.* Such asymmetries may be relatively common; for example Kronforst (2008) reported unidirectional introgression between several pairs of hybridizing *Heliconius* butterflies.

Unlike neutral loci, genes under selection may not be able to move freely across species boundaries, either because they provide local adaptation or result in reduced fitness when in the other species' genetic background. Therefore their introgression rate estimates will be near zero in the two directions. Both *AG-0005F* and *AG-0334P* are accessory gland genes which encode proteins that are transferred to females during mating (Andrés *et al.* in press). Given the reproductive functions of *AG-0005F* and *AG-0334P* it is unlikely that they play a role in adaptations of crickets to local environments (e.g., adaptation to sand versus loam soils (Rand and Harrison 1989; Ross and Harrison 2002,

2006)). More likely they play a role in sperm capacitation, sperm competition, gametic compatibility, and/or male/female interactions. Both of these proteins show several radical amino acid substitutions between species that may contribute to functional differences.

There is a well-documented reproductive incompatibility between *G. firmus* and *G. pennsylvanicus*. Hybrid offspring are only produced from crosses between *G. pennsylvanicus* females and *G. firmus* males, but in the reciprocal cross females produce many fewer eggs, all of which fail to develop (Harrison 1983; Maroja *et al.* chapter 2). In insects many accessory gland proteins have a clear signature of selection (Aguadé 1998, 1999; Begun *et al.* 2000; Swanson *et al.* 2001; Swanson and Vacquier 2002; Andrés *et al.* 2006) and some of these proteins have been shown to influence female oogenesis, ovulation, and oviposition (Wolfner 1997; Neubam and Wolfner 1999; Tram and Wolfner 1999). *AG-0005F* and *AG-0334P* also appear to be under selection, and although their functions are still unknown, their presence in the spermatophore suggests a possible role in the *G. firmus* and *G. pennsylvanicus* reproductive incompatibility. Such a role would explain their lack of introgression.

Using a much smaller sample of populations and individuals, Andrés et al (in press) used a phylogenetic approach and showed that both *AG-0005F* and *AG-0334P* are under selection and have gene genealogies compatible with species exclusivity. Here, with a much larger sample size, we also found a pattern consistent with exclusivity for *AG-0005F* and a paraphyletic pattern for *AG-0334P* (with derived *G. firmus* alleles). Because *G. pennsylvanicus* and *G. firmus* diverged recently and still hybridize, it is expected that most regions of their genome will reveal shared ancestral polymorphism and introgression. Indeed previous efforts to identify diagnostic differences have been unsuccessful (Harrison and Bogdanowicz 1997; Broughton and Harrison 2003). Using a coalescence/population genetics approach, we have also found that most nuclear loci,

including many genes expressed in male accessory gland, remain undifferentiated and exhibit high levels of introgression. The only exceptions are the two genes from accessory gland that are likely under selection. We thus complement the results of Andrés *et al.* (in press) and again reiterate the potential role of *AG-0005F* and *AG-0334P* as barrier genes.

Gene flow and recent demographic history

In contrast to most nuclear loci, mitochondrial DNA did not show evidence of introgression in either direction. This may in part be due to population sampling. Most mixed populations contain southern *G. pennsylvanicus* and southern *G. firmus* haplotypes, but we failed to sample any "pure" *G. pennsylvanicus* populations from the southern part of the range. Because we estimate the extent of introgression using only "pure" populations, our failure to have southern *G. pennsylvanicus* populations represented decreases the probability of detecting introgression. Another factor contributing to the observed lack of introgression for mtDNA is the directionality of the reproductive incompatibility; *m1* is expected to be very low because only *G. pennsylvanicus* females produce hybrids and these hybrids are more likely to backcross to *G. firmus* (Maroja *et al.* unpublished data). Indeed previous studies have found evidence of mtDNA introgression only in the *G. pennsylvanicus* to *G. firmus* direction (Harrison *et al.* 1987; Harrison and Bogdanowicz 1997; Willett *et al.* 1997; Ross and Harrison 2002). Finally the signature of ancestral migration/introgression would be erased more quickly in mtDNA because of its smaller effective population size.

We identified six major mtDNA clades, three of which represent northern *G. firmus* populations and the other three correspond to northern *G. pennsylvanicus,* southern *G. firmus*, and southern *G. pennsylvanicus* populations (Figure 4.2). These are the same mtDNA haplotype groups identified by Willett *et al.* (1997) with two additional clades of northern *G. firmus* individuals (Figure 4.2). As in Willett *et al.* (1997) these clades have

strong support (Figure 4.2) and distinguish *G. firmus* from *G. pennsylvanicus.* However, the mtDNA data still do not provide resolution at the base of the tree and leave unanswered whether each of the species is an exclusive group with respect to mtDNA.

Based on mtDNA haplotype distributions, we also found evidence for a north/south split in both species, again in agreement with Willett *et al.* (1997). With larger sample sizes and broader geographic coverage, it appears that this phylogeographic break runs east-west from the Delmarva Peninsula through northern Maryland and southern Pennsylvania. NBL in southern Pennsylvania contains both clades of *G. pennsylvanicus,* and ESS in northern Maryland contains both clades of *G. firmus.* Although mtDNA clades are geographically well defined, there is evidence of historical or ongoing gene flow between northern and southern *G. pennsylvanicus* populations, e.g., two southern populations (FRN and BRP) include individuals with northern *G. pennsylvanicus* haplotypes (see Figure 4.1). Of the nuclear genealogies, only *EF1-*α showed a pattern consistent with a north/south phylogeographic split (see phylogenetic analyses results and Figures 4.4 - 4.12). The "northern" clade, with bootstrap support of 100%, contains mostly individuals from northern populations. Of the southern populations, only PAR, COV, ESS and FRN had haplotypes in both clades. ESS is located in the mtDNA phylogeographic north/south split and COV and FRN are mixed populations, the latter including northern *G. pennsylvanicus mtDNA* haplotypes.

In agreement with our introgression estimates, the extent of allele sharing between *G. firmus* and *G. pennsylvanicus* at most loci suggests extensive gene flow. The high levels of genetic variation and lack of significant Tajima's *D*, suggests rapid speciation without population bottlenecks. In this scenario ancestral polymorphism would persist even if species barriers were complete (i.e. no hybridization), because only after > 2Ne generations (Maddison 1997) are taxa expected to become reciprocally monophyletic for most loci (Tajima 1983; Neigel and Avise 1986; Harrison 1991; Hey 1994). Given that

these crickets are likely to have large effective population sizes (IMa estimate of over 1 million - see Isolation and introgression Results) and are still exchanging genes, it will be a long time until complete reciprocal monophyly is achieved.

Given the current geographic distribution of the two crickets, we expected to see a signal of population expansion. The northern part of the current range of both species became inhabitable only about 15,000 years ago (Davis 1976; Dyke and Prest 1987), which would suggest that populations must have expanded their numbers recently. However, the lack of a significant Tajima's *D* for most loci indicates that the population expansion was not so substantial as to leave a lasting genetic signature. Of the two species, *G. firmus* has higher average nucleotide diversity, an observation consistent both with a phylogeographic history in which the sizes of *G. firmus* populations may have been greater during past glaciation cycles, because of its association with sandy soils and coastal habitats, and higher introgression from *G. pennsylvanicus* alleles. Our divergence estimate suggests that *G. firmus* and *G. pennsylvanicus* divergence predates the most recent glacial advance. If the rate of mtDNA evolution is 1.2 % per million years per lineage (Brower 1994) using the time since split ($t/\mu=4.29$) calculated with IMa gives an estimative of 202,320 years from a common ancestor, which is in agreement with estimates of Broughton and Harrison (2003) (0.1Ne ~200,000 years) and Willett *et al.* (1997) (187,500 years).

History and structure of the hybrid zone

Based both on morphology and mtDNA phylogeny it appears that the hybrid zone is wider than once thought (Harrison *et al.* 1997). In previous studies the hybrid zone was defined as a long but narrow zone extending from the Blue Ridge Mountains in Virginia to southern Connecticut (Harrison and Arnold 1982; Harrison and Bogdanowicz 1997). However Harrison and Arnold (1982) reported mixed populations in the Shenandoah Valley and speculated that the hybrid zone might also extend to the west of the Blue

Ridge. Indeed we found mixed populations (COV, FRN, MOO) in the Appalachian Mountains west of the Shenandoah Valley. Individuals from the COV, FRN, and MOO populations had substantial variation in color and body size (data not shown) and were on average larger than pure *G. pennsylvanicus*, more similar to *G. firmus*. These populations also included crickets with *mtDNA* haplotypes from both *G. firmus* and *G. pennsylvanicus* clades (Figure4.1 and Figure 4.2).

The zone of overlap between *G. firmus* and *G. pennsylvanicus* is likely a result of secondary contact between previously isolated forms (Willett *et al.* 1997). Because both of these cricket species are inhabitants of grassy fields and disturbed open areas, they have presumably benefited from extensive human habitat alterations and are currently found in large numbers in suburban and even urban areas and along road sides as well as in pastures and other open fields. The increased amount of suitable habitat probably has provided avenues for range expansion and increased gene flow/hybridization between the two species. It is thus possible that the hybrid zone has been expanding.

In ground crickets of the genus *Allenomobius*, Howard and Waring (1991) described a mosaic hybrid zone in which altitude determines the relative abundance of two hybridizing species. A northern species, *Allenomobius fasciatus* and a southern species, *A. socius*, meet in the Appalachians. Along a transect through this region, *A. fasciatus* is most abundant at high elevations whereas *A. socius* predominates at lower elevations (Howard and Waring 1991). The *Gryllus* hybrid zone shows similar features, with the two hybridizing species segregated to some extent by altitude. Outside of the hybrid zone, *G. pennsylvanicus* is found primarily in inland/upland situations, whereas *G. firmus* is costal/lowland (Harrison and Arnold 1982). The hybrid zone along the eastern front of the Blue Ridge occurs along a steep elevational transect. All of the sites that we sampled that are to the west of the Shenandoah Valley occur at relatively low elevations (COV: 354m; MOO: 285m; FRN: 551m), which may explain why these populations are mixed

rather than pure *G. pennsylvanicus*. The Shenandoah Valley might thus have provided a migration route for *G. firmus* individuals to colonize suitable habitats further west, producing a mosaic of pure and mixed populations in the mountain and valley regions of Virginia and West Virginia. This colonization may be quite recent, caused by *G. firmus* moving along roads and/or river drainages.

The expansion of the hybrid zone does not imply that species identities will be eventually erased in a hybrid swarm. As in other insect hybrid zones (e.g., Mendelson and Shaw 2002; Bailey *et al.* 2004), *G. firmus* and *G. pennsylvanicus* have multiple trait differences that restrict gene flow. Some of these barriers operate throughout the zone, whereas others vary geographically. For example, the one-way incompatibility between *G. firmus* females and *G. pennsylvanicus* males has been shown to be characteristic of crickets from both Connecticut and Virginia, whereas a clear soil association has only been documented in Connecticut (Rand and Harrison 1989; Ross and Harrison 2002). Temporal isolation (due to differences in development time) is observed in Virginia but not in Connecticut (Harrison 1985). This barrier may be of particular importance in mixed populations along the Blue Ridge and southern Appalachians, because of the interaction between intrinsic differences in development rate between the species and the variation in length of growing season along elevational gradients. In addition a clear female preference for conspecific males has been demonstrated between Ithaca, NY and Guilford, CT pure species populations, with F1 hybrids behaving like the parental *G. firmus* species (Maroja *et al.* unpublished data). None of these barriers acting alone is complete, but together they appear to severely restrict gene exchange; very few F1 individuals are found in mixed populations and the hybrid zone remains clearly bimodal.

Conclusions

Although independent species will ultimately exhibit divergence across their entire genome, persistence of shared ancestral polymorphism and introgression cause

recently diverged species to be mosaics with respect to genetic differentiation. Depending on the genetic architecture (Ting *et al.* 2001) and as long as alleles at barrier genes do not introgress, species integrity can be maintained even in face of substantial gene flow. Indeed, multilocus studies of closely related species often report discordant genealogical patterns despite well defined boundaries based on morphological, behavioral and ecological characters (Beltran *et al.* 2002; Broughton and Harrison 2003; Machado and Hey 2003; Dopman *et al.* 2005; Putnam *et al.* 2007). In accord with these studies, we report discordant genealogical patterns and differential introgression rates across the genome of the two hybridizing cricket species. The most dramatic outliers are the two accessory gland loci under selection, *AG-0005F* and *AG-0334P,* which showed near zero introgression and more structured species trees. *AG-0005F* and *AG-0334P* are candidate barrier genes with possible reproductive functions in the field crickets *G firmus* and *G. pennsylvanicus.*

ACKNOWLEDGMENTS

We thank Thomas M. McCleskey for writing a program which formats sequence data for PHASE analysis and reincorporates the PHASE output into a nexus format. We thank Steve Bogdanowicz for technical advice in all aspects of this study. Part of this work was carried out by using the resources of the Computational Biology Service Unit (CBSU) at Cornell University, which is partially funded by Microsoft Corporation. Funding was provided by National Science Foundation grants DEB-0111613 (Willie Swanson, PI) and DEB-0639904 to RGH and a AAUW American fellowship to LSM.

REFERENCES

Aguadé, M. 1998. Different forces drive the evolution of the *Acp*26Aa and *Acp*26Ab accessory gland genes in the *Drosophila melanogaster* species. Genetics 150: 1079-1089.

Aguadé, M. 1999. Positive selection drives the evolution of the *Acp*29AB accessory gland protein in *Drosophila*. Genetics 152: 543-551.

Alexander, R.D. 1957. The taxonomy of the field crickets of the eastern United States (Orthoptera: Gryllidae: *Acheta*). Ann. Entomol. Soc. Am. 50: 584-602.

Andrés, J. A., and G. Arnqvist. 2001 Genetic divergence of the seminal signal-receptor system in houseflies: the footprints of sexually antagonistic coevolution? Proc. R. Soc. Lond. B 268: 399-405.

Andrés, J.A., L.S. Maroja, and R.G. Harrison. (in press). Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. Proc. R. Soc. Lond. B.

Andrés, J.A., L.S. Maroja, S.M. Bogdanowicz, W. Swanson, and R.G. Harrison. 2006. Molecular evolution of seminal proteins in field crickets. Mol. Biol. Evol. 23: 1574-1584.

Anisimova M., R. Nielsen and Z. Yang. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164: 1229-1236.

Arnold M.L. 1997. *Natural hybridization and evolution*. New York: Oxford University Press.

Bailey, R.I., C.D. Thomas, and R.K. Butlin. 2004. Premating barriers to gene exchange and their implications for the structure of a mosaic zone between *Chorthippus brunneus* and *C. jacobsi* (Orthoptera: Acrididae). J. Evol. Biol. 17: 108-119

Barbash, D., D. Sinno, A. Tarone, and J. Roote. 2003. A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. Proc. Natl. Acad. Sci. USA 100: 5302-5307.

Barton, N.H. and G.M. Hewitt. 1981. Hybrid zones and speciation. Pp 109-145. in Evolution and Speciation (Atchley, W.R. and D.S. Woodruf, Eds). Cambridge University Press, Cambridge, UK.

Begun, D.J., P. Whitley, B.L. Todd, H.M. Waldrip-Dail, and A.G. Clark. 2000. Molecular population genetics of male accessory gland proteins in *Drosophila*. Genetics 156: 1879-1888.

Beltran M., C.D. Jiggins, V. Bull, M. Linares, J. Mallet, W.O. McMillan, and E. Berminghan. 2002. Phylogenetic discordance at the species boundary: comparative gene genealogies among rapidly radiating *Heliconius* butterflies. Mol. Biol. Evol. 19: 2176-2190.

Besansky N.J., J. Krzywinski, T. Lehmann, F. Simard, M. Kern, O. Mukabayire, D. Fontenille, Y. Touré, N.F. Sagnon. 2003. Semipermeable species boundaries between *Anopheles Gambia* and *Anopheles arabiensis*: evidence from multilocus DNA variation. Proc. Natl. Acad. Sci. USA 100: 10818-10823.

Braswell, W. E., J.A. Andrés, L.S. Maroja, R.H. Harrison, D.J. Howard, and W.J. Swanson. 2006. Identification and comparative analysis of accessory gland proteins in Orthoptera. Genome 84: 1-13.

Brideau N.J., H.A. Flores, J. Wang, S. Maheshware, X. Wang, and D.A. Barbash. 2006. Two Dobzjansky-Muller genes interact to cause hybrid lethality in *Drosophila*. Science 314: 1292- 1295.

Broughton, R.E. and R.G. Harrison. 2003. Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. Genetics 163: 1389-1401.

Brower, A.V.Z. 1994. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. Proc. Natl. Acad. Sci. USA 91: 6491-6495.

Clark, N. L., J.E. Aagaard, and W.J. Swanson. 2006 Evolution of reproductive proteins from animals and plants. J. Reprod. Fert. 131: 11-22.

Coyne, J. A. and H.A. Orr. 2004. *Speciation*. Sunderland, MA: Sinauer Associates.

Davis, M.B. 1976. Pleistocene biogeography of temperate deciduous forests. Geosci and Man, 13: 13-26.

Doherty, J.A. and M. Storz. 1992. Calling song and selective phonotaxis in field crickets, *Gryllus firmus* and *G. pennsylvanicus* (Orthoptera: Gryllidae). J. Insect Behav. 5: 555-569.

Dopman, E. B., L. Peréz, S.M. Bogdanowicz, and R.G. Harrison. 2005 Consequences of reproductive barriers for genealogical discordance in the European corn borer. Proc. Natl. Acad. Sci. 102: 14706-14711.

Dyke, A.S. and V.K. Prest. 1987. Late Wisconsinan and Holocene history of the Laurentide ice sheet. Geogr. Phys. Quaternaire 41: 237-264.

Excoffier, L.P., P.E. Smouse, and J.M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. Genetics 131: 479-491.

Farris J. S., M. Källersjö, A.G. Kluge, and C. Bult. 1995. Constructing a significance test for incongruence. Syst. Biol. 44: 570–572.

Goldman, N. and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. 11: 725-736.

Grahame J.W., C.S. Wilding and R.K. Butlin. 2006. Adaptation to a steep environmental gradient and an associated barrier to gene exchange in *Littorina saxatilis.* Evolution 60: 268-278.

Grant V. 1981. *Plant Speciation*. New York, Columbia University Press.

Grant, P.R., B.R. Grant, J.A. Markert, L.F. Keller, and K. Petren. 2004. Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. Evolution 58: 1588-1599.

Harrison, R.G. 1979. Speciation in North American field crickets: evidence from electrophoretic comparisons. Evolution 33: 1009- 1023.

Harrison, R.G. 1983. Barriers to gene exchange between closely related cricket species. I. Laboratory hybridization studies. Evolution 37: 245-251.

Harrison, R.G. 1985. Barriers to gene exchange between closely related cricket species. II. Life cycle variation and temporal isolation. Evolution 39: 244-259.

Harrison, R.G. 1990. Hybrid zones: windows on evolutionary processes. Pp: 69-128 *in* Oxford Surveys in Evolutionary Biology Vol 7 (Futuyma, D. and J. Antonovics, Eds). Oxford University Press, New York.

Harrison, R.G. and D.M. Rand. 1989. Mosaic hybrid zones and the nature of species boundaries. Pp 111-133 *in* Speciation and its consequences (Otte D., and J.A. Endler, Eds). Sinauer, Sunderland, MA.

Harrison, R.G. and J. Arnold. 1982. A narrow hybrid zone between closely related cricket species. Evolution 36: 355-552.

Harrison, R.G. and S.M. Bogdanowicz. 1997. Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. Evolution 51: 493-505.

Harrison, R.G., D.M. Rand, and W.C. Wheeler. 1987. Mitochondrial DNA variation in field crickets across a narrow hybrid zone. Mol. Biol. Evol. 4: 144-158.

Harshman, L.G., and T. Prout. 1994. Sperm displacement without sperm transfer in *Drosophila melanogaster*. Evolution 48: 758-766.

Herndon, L.A., and M.F. Wolfner. 1995. A *Drosophila* seminal protein, *Acp*26Aa, stimulates egg laying in females for 1 day after mating. Proc Natl. Acad. Sci. USA 92: 10114-10118.

Hey, J. 1994. Bridging phylogenetics and population genetics with gene tree models, pp. 435-449. in *Molecular Ecology and Evolution: Approaches and Applications*, edited by B. Schierwater, B. Streit, G. Wagner and R. DeSalle. Birkhauser, Boston.

Hey, J. 2005. On the number of New World founders: a population genetic portrait of the peopling of the Americas. PLoS Biol. 3: 965–975.

Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence if *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167: 747-760.

Hey, J., and R. Nielsen. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc Natl. Acad. Sci. USA 104: 2785–2790.

Howard, D.J., and G.L. Waring. 1991. Topographic diversity, zone width and the strength of reproductive isolation in a zone of overlap and hybridization. *Evolution* 45: 1120-1135.

Hudson, R.R. 1983. Properties of a neutral allele model with intragenic recombination Theor. Popul. Biol. 23, 183-201.

Hudson, R.R. 1992. Gene trees, species trees and the segregation of ancestral alleles. Genetics 131: 509-512.

Hudson, R.R., and J.A. Coyne. 2002. Mathematical consequences of the genealogical species concept. Evolution 56: 1557-1565.

Huelsenbeck J.P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754-755.

Jain R., M.C. Rivera, J.E Moore, and J.A. Lake. 2002. Horizontal gene transfer in microbial genome evolution. Theor. Popul. Biol. 61: 489-495.

Kronforst, M.R. 2008. Gene flow persists millions of years after speciation in *Heliconius* butterflies. BMC Evolutionary Biology 8: 98.

Kuhner M.K. and L.P Smith. 2007. Comparing likelihood and Bayesian coalescent estimation of population parameters. Genetics 175: 155–165.

Li, N., and M. Stephens. 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165: 2213-2233.

Machado, C.A., and J. Hey. 2003 The causes of phylogenetic conflict in a classic *Drosophila* species group. Proc. R. Soc. Lond. B 270: 1193-1202.

Machado, C.A., R.M. Kilman, J.A. Markert, and J. Hey. 2002. Inferring the history of Speciation from multilocus DNA sequence data: The case of *Drosophila pseudoobscura* and close relatives. Mol. Biol. Evol. 19: 472-488.

Maddison, W.P. 1997. Gene trees in species trees. Syst. Biol. 46: 523-536.

Maroja, L.S., J.A. Andrés, and R.G. Harrison. (chapter 2). Multiple barriers to gene exchange in the *Gryllus firmus* - *Gryllus pennsylvanicus* hybrid zone.

Maroja, L.S., M. E. Clark and R.G. Harrison. (chapter 3). *Wolbachia* plays no role in the one-way reproductive incompatibility between the hybridizing field crickets *Gryllus firmus* and *G. pennsylvanicus*.

Masly J.P., C.D. Jones, M.A.F. Noor, and H.A. Orr. 2006. Gene transposition as a cause of hybrid sterility. Science 313: 1448-1450.

Mayr E. 1963. *Animal Species and Evolution*. Harvard University Press, Cambridge, MA.

Mendelson, T.C. and K.L. Shaw. 2002. Genetic and behavioral components of the cryptic species boundary between *Laupala cerasina* and *L. kohalensis* (Orthoptera: Gryllidae). Genetica 116: 301-310.

Neigel, J. E. and J. C. Avise. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. Pp 515-534 *in* Evolutionary Processes and Theory (S. Karlin and E. Nevo, Eds). Academic Press, New York.

Neubaum, D. M., and M. F. Wolfner. 1999. Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. Genetics 153: 845-857.

Nichols, R. 2001. Gene trees and species trees are not the same. Trend Ecol. Evol. 16: 358-364.

Nielsen, R. and J. Wakeley. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics 158: 885-896.

Noor, M. and J.L. Feder. 2006. Speciation genetics: evolving approaches. Nat. Rev. Genet. 7: 851-861.

Nosil, P., S.P. Egan, and D.J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. Evolution 62: 316-336.

Payseur, B., and M. Nachman. 2005. The genomics of speciation: investigation investigating the molecular correlates of X chromosome introgression across the hybrid zone between *Mus domesticus* and *Mus musculus*. Biol. J. Linn. Soc 84: 523-534.

Posada, D. and K.A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817-818.

Presgraves, D., L. Balagopalan, S. Abmayr, and H. Orr. 2003. Adaptive evolution drives divergence of hybrid incompatibility gene between two species of *Drosophila*. Nature 243: 715-719.

Putman, A. S., J.M. Scriber, and P. Andolfatto. 2007. Dicordant divergence times among z-chromosome regions between two ecologically distinct swallowtail butterfly species. Evolution 61: 912-927.

Rand, D.M. and R.G. Harrison. 1989. Ecological genetics of a mosaic hybrid zone: mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. Evolution 43: 432-449.

Rieseberg, L. 1997. Hybrid origins of plant species. Annu. Rev. Ecol. Syst. 28: 359-389.

Rieseberg, L.H., J. Whitton, and K. Gardner. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics 152: 713-727.

Ross, C.L. and R.G. Harrison. 2002. A fine-scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus*. Evolution 56: 2296-2312.

Ross, C.L. and R.G. Harrison. 2006. Viability selection on overwintering eggs in a field cricket mosaic hybrid zone. Oikos 115: 53-68.

Rozas, J., J.C. Sánchez-Delbarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496-2497.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin: a software for population genetics data analysis. Genetics and Biometry Lab, Department of Anthropology, University of Geneva.

Seehausen O. 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19: 198-207.

Shierup, M.H., and J. Hein. 2000. Consequences of recombination on traditional phylogenetic analysis. Genetics 156: 879-891.

Shriner, D., D.C. Nickle, M.A. Jensen, and J.I. Mullins. 2003. Potential impact of recombination on sitewise approaches for detecting positive natural selection. Genet. Res. 81: 115-121.

Stephens, M., N.J. Smith, and P. Donnelly. 2001 A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68: 978-989.

Swanson, W. J., A. Wong, M.F. Wolfner, and C.F. Aquadro. 2004. Evolutionary expressed tag analysis of *Drosophila* Identifies genes subjected to positive selection. Genetics 168: 1457-1465.

Swanson, W. J., A.G. Clark, H.M. Waldrip-Dail, M.F. Wolfner, and C. F. Aquadro. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc. Natl. Acad. Sci. USA 95: 4051-4054.

Swanson, W.J., and V.D. Vacquier. 2002. The rapid evolution of reproductive proteins. Nat. Rev. Genet. 3:137–144.

Swofford, D.L. 2003. *PAUP* 4.0B10* Sunderland, MA: Sinauer.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105: 437-460.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585-595.

Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10: 512–526.

Templeton, A.R. 1994. The role of molecular genetics in speciation studies. in *Molecular Ecology and Evolution: Approaches and Applications*, edited by B. Schierwater, B. Streit, G. Wagner and R. DeSalle. Birkhauser, Boston.

Ting, C., S.-C. Tsaur, and C.-I. Wu. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. Science 282: 1501-1504.

Ting, C.-T., S.-C. Tsaur, and C.-I. Wu. 2000. The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus.* Proc. Natl. Acad. Sci. USA. 97: 5313-5316.

Ting, C-T., A. Takahashi, and C-I Wu. 2001. Incipient speciation by sexual isolation in *Drosophila*: concurrent evolution at multiple loci. Proc. Natl. Acad. Sci. USA 98: 6709-6713.

Tram, U., and M.F. Wolner. 1999. Male fluid proteins are essential for sperm storage in *Drosophila melanogaster*. Genetics 153: 837-844.

Vasemagi, A., J. Nilsson, and C.R. Primmer. 2005. Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). Mol. Biol. Evol. 22: 1067-1076.

Wang, R.L., J. Wakeley, and J. Hey. 1997. Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. Genetics 147: 1091-1106.

Willett, C., M.J. Ford, and R.G. Harrison. 1997. Inferences about the origin of a field cricket hybrid zone from a mitochondrial DNA phylogeny. Heredity 79: 484-494.

Wilson, D.J., and G. McVean. 2006. Estimating diversifying selection and functional constraint in the presence of recombination. Genetics 172: 1411-1425.

Wittbrodt J., D. Adam, B. Malitscheck, W. Maueler, F. Raulf, A. Telling, S.M. Robertson, and M. Schartl. 1989. Novel putative receptor tyrosine kinase encoded by melanoma-inducing *Tu* locus in *Xiphophorus.* Nature 341: 415-421.

Woerner, A.E., P.C. Murray, and M.F. Hammer. 2007. Recombination-filtered genomic datasets by information maximization. Bioinformatics: 23: 1851-1853.

Wolfner, M.F. 1997. Tokens of love: Functions and regulations of *Drosophila* accessory male products. Insect Biochem. Mol. Biol. 27: 179-192.

Wu, C.-I. 2001. The genic view of the process of speciation. J. Evol. Biol. 14: 851-865.

Wu, C.-I., and C-T. Ting. 2004. Genes and speciation. Nat. Rev. Genet. 5: 114-122.